TRƯỜNG ĐẠI HỌC SÂN KHẤU – ĐIỆN ẢNH HÀ NỘI
KHOA CÔNG NGHỆ ĐIỆN ẢNH – TRUYỀN HÌNH

KỸ THUẬT AUDIO – VIDEO

CHƯƠNG 8 · PHẦN C

Nguyên lý nén video

Vì sao phải nén

- Tín hiệu chưa nén vượt xa dung lượng của mọi kênh truyền và thiết bị lưu trữ thông thường
- Nén khả thi vì tín hiệu hình chứa bốn loại dư thừa
- Nén là bỏ đi phần mà bộ giải mã đoán được; cái giá là tính toán nhiều hơn

Tốc độ bit thô (8.19)

Tốc độ bit của tín hiệu chưa nén là tích năm thừa số.

$$R_{\text{raw}} = W \times H \times f_{\text{fr}} \times n \times s$$

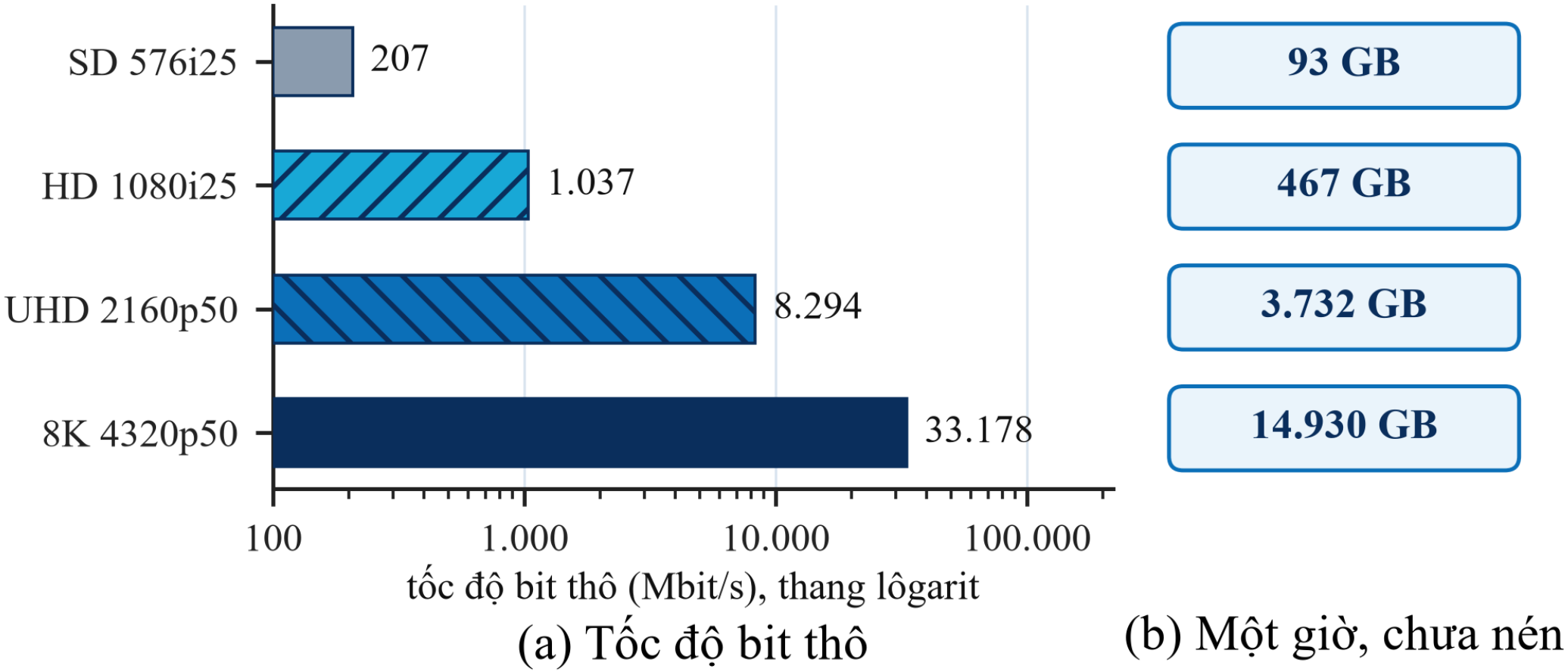
- W và H là số mẫu theo hai chiều, f là tần số hình, n là độ sâu bit
- s là số mẫu trung bình mỗi điểm ảnh theo cấu trúc lấy mẫu con màu (đúng với đại lượng đã tính ở mục 8.9)
- Câu hỏi vì sao phải nén được trả lời bằng chính con số này

Tốc độ bit thô (Bảng 8.27; 10 bit, 4:2:2)

Định dạng	Phép tính	Tốc độ bit	Một giờ nặng
SD 576i25	720 × 576 × 25 × 10 × 2	207 Mbit/s	93 GB
HD 1080i25	1920 × 1080 × 25 × 10 × 2	1 037 Mbit/s	467 GB
UHD 2160p50	3840 × 2160 × 50 × 10 × 2	8 294 Mbit/s	3 732 GB
8K 4320p50	7680 × 4320 × 50 × 10 × 2	33 178 Mbit/s	14 930 GB

- Hơn một gigabit mỗi giây cho một tín hiệu HD; kênh phát sóng số mặt đất ở Việt Nam chỉ có vài megabit mỗi giây cho mỗi chương trình
- Chênh lệch lên tới hàng trăm lần: toàn bộ Phần C và Phần D trình bày cách bù đắp

Tốc độ bit thô và dung lượng một giờ



Hình 8.30 — Tốc độ bit thô và dung lượng một giờ của bốn định dạng

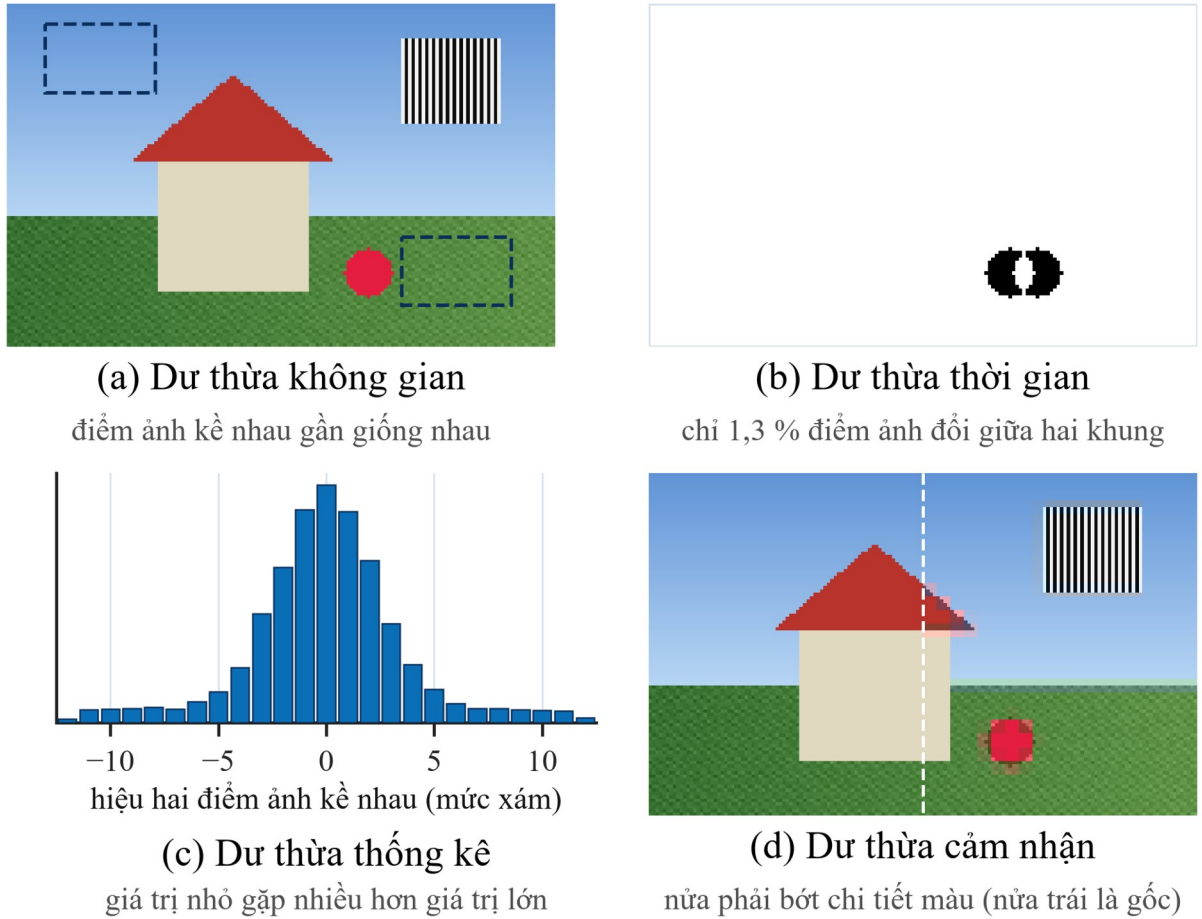
Tính tốc độ bit của tín hiệu thành phần số

GHI NHỚ

864 × 625 × 25 = 13,5 MHz; nhân với 10 bit và với 2 cho 270 Mbit/s, đúng bằng tốc độ SD-SDI.

- Tín hiệu SD có 864 × 625 điểm ảnh mỗi khung (kể cả khoảng xoá), 25 khung/s, 10 bit mỗi mẫu
- 4:2:2 có trung bình hai mẫu mỗi điểm ảnh: một mẫu chói cộng nửa mẫu Cb và nửa mẫu Cr; chỉ 4:4:4 mới có ba mẫu
- Với tám bit, kết quả tương ứng là 216 Mbit/s

Bốn loại dư thừa mà nén khai thác



Hình 8.31 — Bốn loại dư thừa mà kỹ thuật nén khai thác

Bốn loại dư thừa và cách khai thác (Bảng 8.28)

Loại	Bản chất	Khai thác bằng	Tổn hao?
① Không gian	Điểm ảnh cạnh nhau trong một khung gần giống nhau	Biến đổi cosin rời rạc, dự đoán trong khung	Không, nếu chỉ biến đổi
② Thời gian	Hai khung liên tiếp gần như trùng nhau; thường dưới 5 % điểm ảnh thật sự đổi	Bù chuyển động, khung P và B	Không, về nguyên tắc
③ Thống kê	Giá trị nhỏ xuất hiện nhiều hơn giá trị lớn	Mã chạy dài, mã độ dài thay đổi, mã số học	KHÔNG
④ Cảm nhận	Mắt không thấy một phần thông tin có trong tín hiệu	Lấy mẫu con màu, lượng tử hoá có trọng số thị giác	CÓ — chỗ mất mát thật sự

Ranh giới quan trọng nhất của cả Phần C

GHI NHỚ

Ba loại dư thừa đầu là dư thừa thực sự; loại thứ tư đánh đổi thông tin lấy dung lượng.

- Ba loại đầu: loại bỏ vẫn khôi phục nguyên vẹn từng bit — nén không tổn hao
- Loại thứ tư dựa vào độ nhạy kém của mắt người — nén có tổn hao
- Mọi codec video thực dụng khai thác cả bốn loại; vấn đề là mức tổn hao bao nhiêu thì còn chấp nhận được

Nguyên lý đầu tiên: bỏ phần mà bộ giải mã đoán được

- Tín hiệu hình gồm phần **dự đoán được** và phần **bất ngờ**
- Chỉ phần bất ngờ mang thông tin; gửi phần dự đoán được là lãng phí
- Nén gồm hai việc: (1) dự đoán từ dữ liệu bộ giải mã cũng có; (2) mã hoá sai số, tức hiệu giữa bản thật và bản đoán, ở dạng gọn nhất
- Dự đoán tốt thì hiệu hầu hết gần bằng 0, và giá trị thường gặp được ghi bằng ít bit

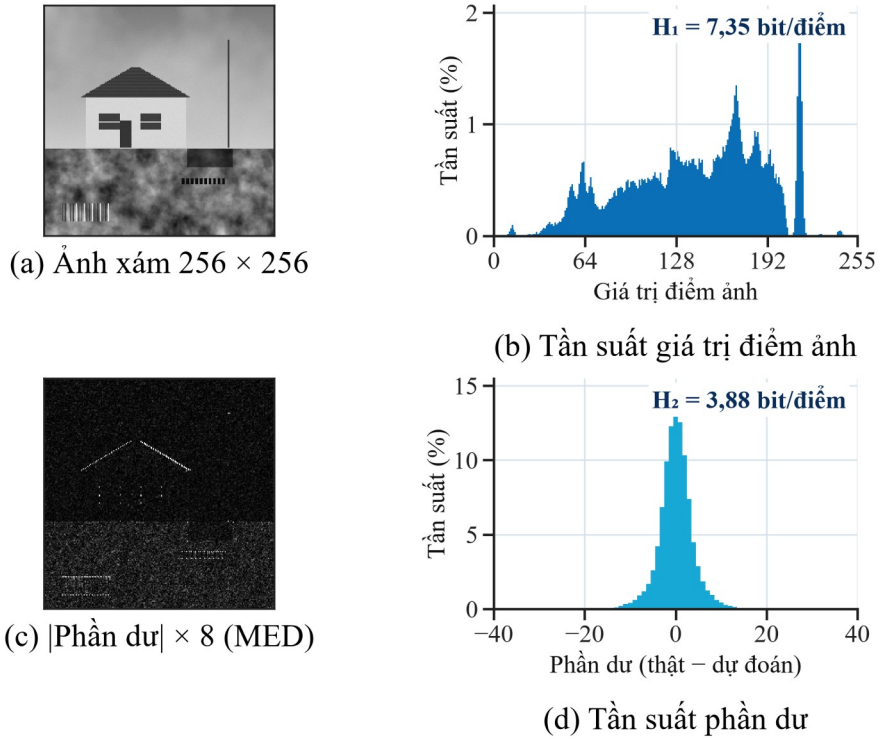
Entropy theo Shannon (8.20)

Mức gọn của sai số dự đoán đo bằng entropy.

$$H = -\sum_i p_i \log_2 p_i \quad \text{(bit)}$$

- p_i là xác suất của giá trị thứ i; H là số bit trung bình tối thiểu để ghi mỗi mẫu bằng bộ mã không tổn hao [27]
- Giá trị hiếm (p nhỏ) là bất ngờ lớn, tốn nhiều bit
- Giá trị hay gặp (p gần 1) hầu như không tốn bit

Dự đoán làm entropy giảm



(a) Ảnh xám 256 × 256

(b) Tần suất giá trị điểm ảnh

(c) |Phần dư| × 8 (MED)

(d) Tần suất phần dư

Hình 8.32 — Biểu đồ giá trị điểm ảnh trải rộng; biểu đồ sai số sau dự đoán co về một đỉnh nhọn quanh 0

- Entropy của ảnh mô phỏng: 7,35 bit/điểm; của sai số dự đoán: 3,88 bit/điểm, bằng 53 % lúc đầu
- Không mất thông tin: bộ giải mã dựng lại bản đoán rồi cộng sai số là có nguyên điểm ảnh

Bốn việc của mọi bộ nén video

GHI NHỚ

Dự đoán — biến đổi — lượng tử hoá — mã hoá entropy.

- Dự đoán từ những gì đã có; biến đổi phần dư để dồn năng lượng về vài hệ số
- Lượng tử hoá loại bỏ phần mắt không cần: khâu DUY NHẤT có tổn hao
- Mã hoá entropy ghi gọn phần còn lại
- JPEG, H.264, HEVC, AV1 giống nhau ở thứ tự; khác nhau ở mức tinh vi khi thực hiện từng việc

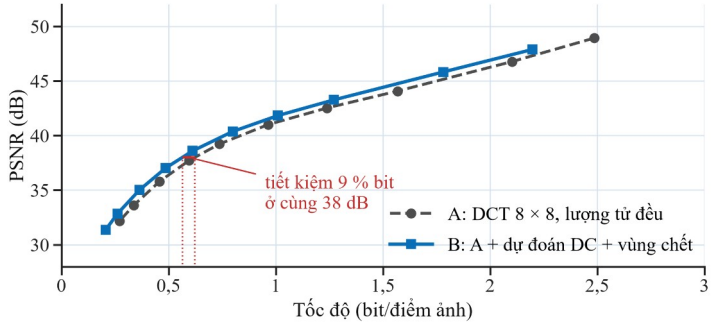
Cái giá của nén kỹ: hàm giá tốc độ–méo (8.21)

Mỗi khối được quyết định theo một hàm giá.

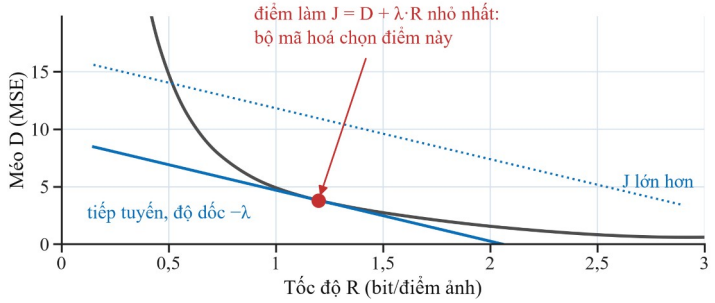
$$J = D + \lambda R$$

- Bộ mã hoá thử từng phương án, đo số bit R và độ méo D, giữ phương án cho J nhỏ nhất
- D thường là tổng bình phương sai khác so với bản gốc
- λ càng lớn thì càng ưu tiên tiết kiệm bit hơn giữ chất lượng

Đường tốc độ–méo và số phương án phải thử



(a) Hai bộ mã hoá, cùng ảnh, đo thật



(b) Đường D–R của A và tiếp tuyến

Hình 8.33 — Công cụ tốt hơn dịch đường sang trái; bộ mã hoá chọn điểm làm J = D + λR nhỏ nhất

- Ở 38 dB, công cụ tốt hơn tiết kiệm khoảng 9 % bit trong thí nghiệm này
- Số cách chia cây tứ phân khối 64 × 64 xuống 8 × 8: f(0) = 1, f(k) = 1 + f(k−1)^4 ⇒ f(1) = 2, f(2) = 17, f(3) = 83 522

Ba thế hệ codec làm tinh bốn việc (Bảng 8.29)

Công cụ	H.264 / AVC (2003)	H.265 / HEVC (2013)	H.266 / VVC (2020)
Khối mã hoá lớn nhất	16 × 16	tới 64 × 64 (cây tứ phân)	tới 128 × 128
Hướng dự đoán trong khung	9	35	67
Cỡ biến đổi	4 × 4, 8 × 8	4 × 4 tới 32 × 32	tới 64 × 64
Mã hoá entropy	CAVLC hoặc CABAC	CABAC	CABAC cải tiến
Lọc trong vòng lặp	lọc khối	lọc khối + SAO	thêm ALF
Tiết kiệm bit so với thế hệ trước	—	khoảng 40 – 50 % [29]	khoảng 40 – 50 % [30]
Chi phí mã hoá	—	gấp nhiều lần	gấp nhiều lần nữa

Những điều cần nhớ

- Tín hiệu HD chưa nén cần hơn 1 Gbit/s; kênh phát sóng chỉ có vài Mbit/s
- Bốn loại dư thừa: không gian, thời gian, thống kê, cảm nhận; chỉ loại cảm nhận gây mất mát
- Nén = dự đoán + mã hoá sai số; entropy đo mức gọn tối thiểu
- Bốn việc: dự đoán, biến đổi, lượng tử hoá, mã hoá entropy
- Nén kỹ hơn thì bộ mã hoá phải thử nhiều phương án hơn ($J = D + \lambda R$)

Nén không tổn hao và nén có tổn hao

- Hai định nghĩa chặt và hai mức tỉ số nén rất khác nhau
- Đo tổn hao bằng máy (PSNR, SSIM, VMAF) và bằng người (ITU-R BT.500)
- Tổn hao thế hệ khi mã hoá lại; nơi vẫn cần nén không tổn hao

Hai định nghĩa chặt

Nén không tổn hao

- Giải nén cho dữ liệu giống bản gốc từng bit
- Tỉ số nén với ảnh thật thường chỉ 1,5 – 2 lần
- Không cố định được tốc độ bit vì tỉ số phụ thuộc nội dung; cảnh nhiều chi tiết gần như không nén được

Nén có tổn hao

- Dữ liệu sau giải nén khác bản gốc, ở những chỗ mắt ít nhận ra
- Tỉ số nén hàng chục tới hàng trăm lần
- Tốc độ bit có thể đặt theo yêu cầu

Tỉ số nén và dung lượng tệp (8.22)

Hai đại lượng dùng thường xuyên khi đặt thông số mã hoá.

$$\mathrm{CR} = \frac{R_{\mathrm{raw}}}{R_{\mathrm{enc}}} \; ; \qquad S = \frac{R_{\mathrm{enc}}\, t}{8}$$

- CR là tỉ số nén; R(thô) và R(nén) là tốc độ bit trước và sau nén
- S là dung lượng tệp (byte) khi R(nén) tính bằng bit/s và t tính bằng giây; chia 8 để đổi bit ra byte

Digital Betacam là nén có tổn hao nhẹ

GHI NHỚ

Digital Betacam: 128 Mbit/s, tỉ số nén khoảng 2:1 so với 270 Mbit/s (khoảng 1,6:1 so với 207 Mbit/s hình tích cực), dùng DCT có tổn hao rất nhẹ.

- Xếp vào loại "gần như không nhìn thấy khác biệt", không thuộc nhóm không tổn hao
- Phép kiểm tra đơn giản: nén không tổn hao không bao giờ cố định được tốc độ bit
- Định dạng ghi băng cam kết tốc độ bit không đổi chắc chắn có tổn hao, vì nếu không thì cảnh phức tạp sẽ không ghi vừa băng

Đo bằng máy: sai số bình phương trung bình (8.23)

Trung bình bình phương sai khác từng điểm ảnh so với bản gốc.

$$\mathrm{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left[I(i,j) - \hat{I}(i,j) \right]^2$$

- I là ảnh gốc, Î là ảnh sau giải nén, kích thước M × N
- MSE càng nhỏ thì ảnh càng gần bản gốc

Đo bằng máy: PSNR (8.24)

Tỉ số tín hiệu cực đại trên nhiễu của ảnh nén.

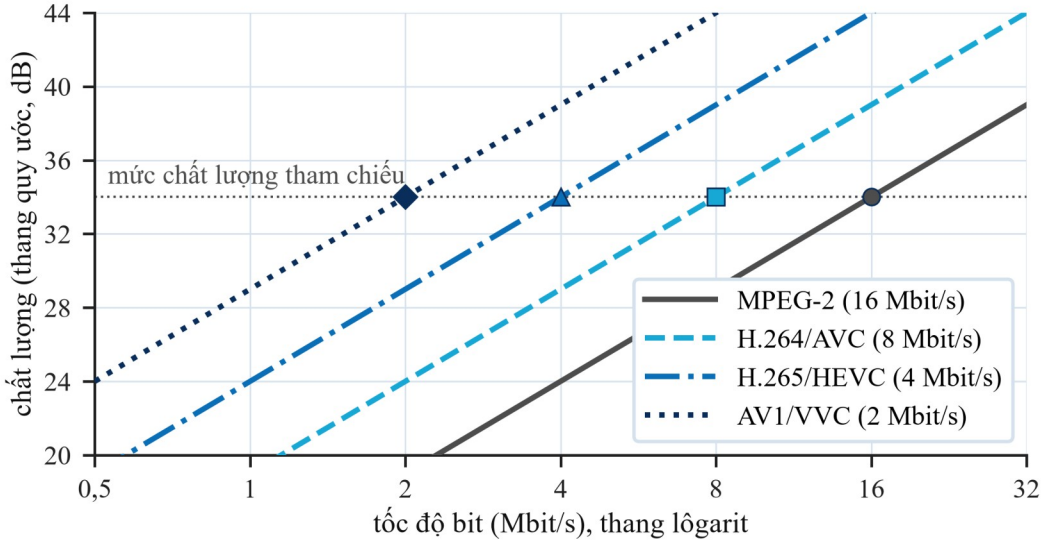
$$\mathrm{PSNR} = 10\log_{10}\frac{(2^n-1)^2}{\mathrm{MSE}} \quad (\mathrm{dB})$$

- n là độ sâu bit; tử số là bình phương giá trị lớn nhất của một mẫu
- PSNR càng cao thì sai khác càng nhỏ, nhưng không khớp hoàn toàn với cảm nhận

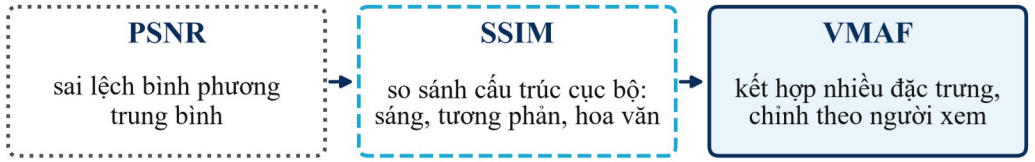
Ba chỉ số đo chất lượng bằng máy (Bảng 8.30)

Chỉ số	Cách đo	Ưu điểm	Nhược điểm
PSNR	Từ sai lệch bình phương trung bình so với ảnh gốc	Rất dễ tính, có mặt trong mọi công cụ	KHÔNG khớp cảm nhận: ảnh mờ đều thường điểm cao hơn ảnh sắc nét có ít nhiễu
SSIM	So sánh cấu trúc cục bộ: độ sáng, tương phản, hoa văn	Gần mắt người hơn PSNR đáng kể	Vẫn không tính chuyển động và điều kiện xem
VMAF	Kết hợp nhiều đặc trưng, hiệu chỉnh theo điểm chấm của người xem thật	Bám sát cảm nhận nhất; nền tảng phát trực tuyến dùng nhiều nhất	Tốn công tính; phụ thuộc mô hình đã huấn luyện

Đường chất lượng – tốc độ bit của bốn thế hệ codec



(a) Mỗi thế hệ đạt cùng chất lượng ở khoảng một nửa tốc độ bit



(b) Ba chỉ số đo chất lượng bằng máy

Hình 8.34 — Đường chất lượng – tốc độ bit của bốn thế hệ codec, và ba chỉ số đo chất lượng

Quy tắc ba lần

QUY TẮC

Mỗi lần mã hoá lại một tệp đã nén có tổn hao đều làm mất thêm chất lượng, dù tốc độ bit không giảm.

- Ít nhất ba lần mã hoá trong dây chuyền thông thường: máy quay ghi thẻ, xuất bản dựng, nền tảng phát hành
- Bộ mã hoá lần hai phải mã hoá cả lỗi mà lần một đã sinh ra
- Dựng trên codec chỉ nén trong khung (mục 8.24); chỉ nén sâu ở khâu cuối cùng

Ba trường hợp dùng nén không tổn hao (Bảng 8.31)

Trường hợp	Vì sao cần	Định dạng thường dùng
Lưu trữ tư liệu gốc của viện phim, thư viện	Phải khôi phục nguyên vẹn; tổn hao tích luỹ qua các lần chuyển đổi	FFV1 trong Matroska: có tài liệu công khai, không ràng buộc bản quyền
Ảnh tĩnh và đồ hoạ trong sản xuất	Chữ và đường nét sắc; nén có tổn hao làm nhoè viền ngay ở mức nhẹ	PNG; TIFF không nén hoặc LZW
Bản trung gian giữa hai công đoạn hậu kỳ	Tránh cộng dồn tổn hao thế hệ	ProRes 4444 XQ, DNxHR 444 (danh nghĩa có tổn hao nhưng không đo được trong thực tế)

- Cả ba thuộc khâu lưu trữ và trung gian, không phải khâu phát hành
- Đến tay khán giả, nén có tổn hao luôn có ưu thế vì tỉ số nén hàng chục lần không đạt được bằng cách nào khác

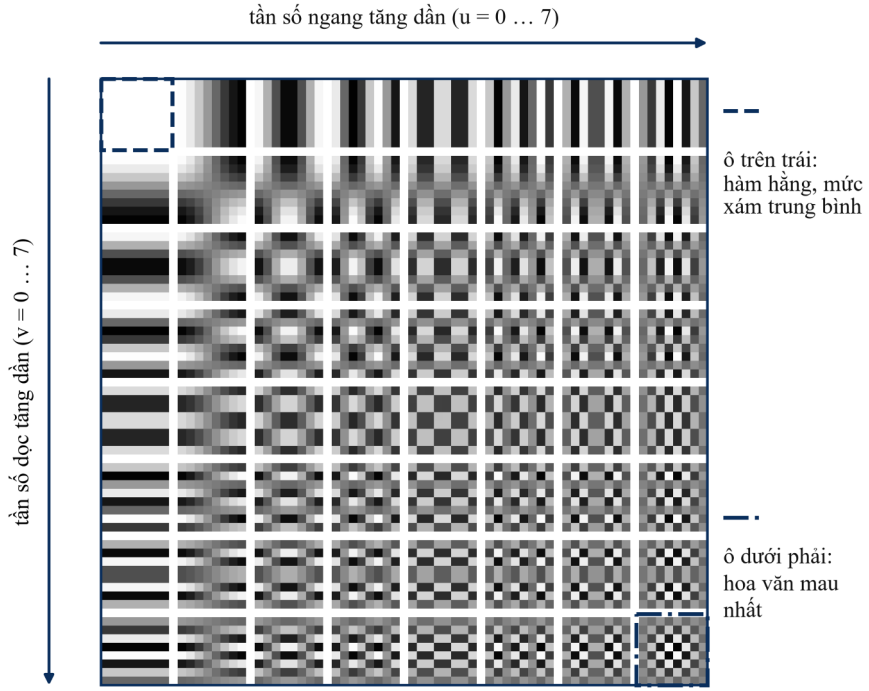
Những điều cần nhớ

- Không tổn hao: giống bản gốc từng bit, chỉ 1,5 – 2 lần, không cố định được tốc độ bit
- Có tổn hao: hàng chục tới hàng trăm lần, đặt được tốc độ bit
- PSNR dễ tính nhưng không khớp cảm nhận; VMAF bám sát nhất; BT.500 là căn cứ cuối cùng
- Mỗi lần mã hoá lại là mất thêm chất lượng (tổn hao thế hệ)
- Nén không tổn hao dùng ở lưu trữ, đồ hoạ và bản trung gian

Nén trong khung

- Nén từng ảnh độc lập, không dùng ảnh lân cận
- Chuỗi bước: biến đổi cosin, lượng tử hoá, quét zig-zag, mã hoá entropy
- Nhận ra vỡ khối, nhiễu muỗi và bệt chi tiết bằng mắt

Vì sao đi qua miền tần số



(a) 64 hàm cơ sở của biến đổi DCT 8 × 8

Hình 8.35 — 64 hàm cơ sở của biến đổi cosin rời rạc 8 × 8

- Trong miền điểm ảnh mọi giá trị quan trọng như nhau; sau biến đổi, phần quan trọng dồn về góc trên trái
- Phần mắt ít nhạy tập trung ở góc dưới phải: có thể loại bỏ có chọn lọc

Biến đổi cosin rời rạc 8 × 8 (8.25)

Khối 8 × 8 điểm ảnh được viết lại thành 64 hệ số.

$$F(u,v) = \frac{1}{4} C_u C_v \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16}$$

- f(x,y) là giá trị điểm ảnh, F(u,v) là hệ số tần số; F(0,0) là hệ số DC
- Hệ số Cw theo (8.26): Cw = 1/√2 khi w = 0 và Cw = 1 khi w > 0

Biến đổi ngược (8.27)

Phép biến đổi ngược dùng cùng các hàm cơ sở.

$$f(x,y)=\frac{1}{4}\sum_{u=0}^{7}\sum_{v=0}^{7}C_uC_vF(u,v)\cos\frac{(2x+1)u\pi}{16}\cos\frac{(2y+1)v\pi}{16}$$

- Đưa hệ số về điểm ảnh
- Nếu không có lượng tử hoá xen giữa, khôi phục chính xác từng giá trị

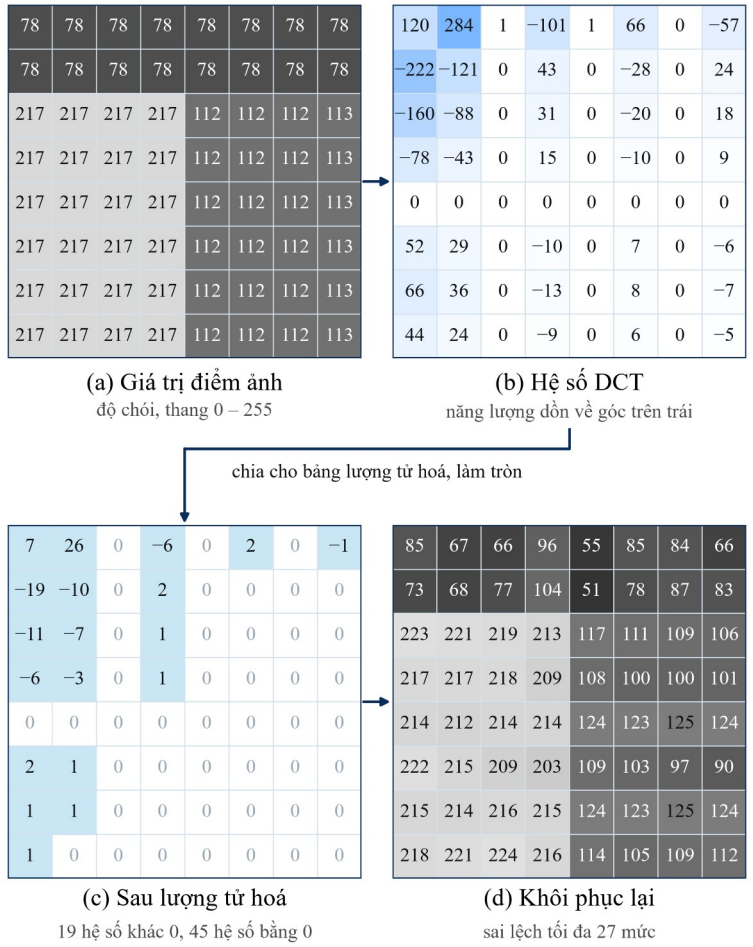
Điều cần nhớ về DCT

GHI NHỚ

Biến đổi cosin rời rạc không làm mất thông tin nào.

- Nó chỉ viết lại cùng một khối 64 số dưới dạng khác
- Toàn bộ mất mát của nén trong khung xảy ra ở bước sau: lượng tử hoá các hệ số
- Nắm được điều này phân biệt người hiểu bản chất với người chỉ nhớ chuỗi bước

Bốn bước trên một khối 8 × 8 thật



Hình 8.36 — Bốn bước của nén trong khung trên một khối 8 × 8 thật

Lượng tử hoá (8.28)

Chia từng hệ số cho một số rồi làm tròn: bước duy nhất gây tổn hao.

$$F_q(u,v) = \text{round}\left(\frac{F(u,v)}{q \cdot Q(u,v)}\right)$$

- Q(u,v) là bảng lượng tử hoá; q là hệ số nhân (thanh chất lượng)
- Sau bước này, 45 trong 64 hệ số trở thành 0: chỉ cần ghi 19 giá trị thay vì 64
- Sai lệch tối đa của khối khôi phục chỉ vài chục mức trên thang 256, tập trung ở ranh giới sắc nét, nơi mắt ít chú ý

Bảng lượng tử hoá: kiến thức thị giác đi vào kỹ thuật

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

(a) Bảng lượng tử hoá JPEG

15	52	0	−13	0	3	0	−2
−37	−20	0	5	0	−1	0	1
−23	−13	0	3	0	−1	0	1
−11	−5	0	1	0	0	0	0
0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0
3	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0

(b) Bảng × 0,5: chất lượng cao

2	6	0	−2	0	0	0	0
−5	−3	0	1	0	0	0	0
−3	−2	0	0	0	0	0	0
−1	−1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(c) Bảng × 4: chất lượng thấp

bảng nhân	số hệ số khác 0	sai lệch tối đa
× 0,5	24	11 mức
× 1	19	27 mức
× 4	11	55 mức

hàng × 1 chính là khối ở Hình 8.36

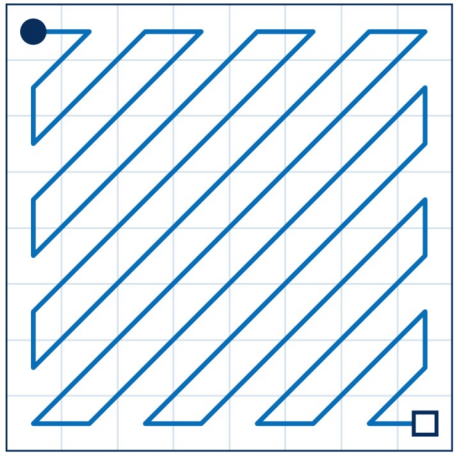
(d) Tóm tắt ba mức

Hình 8.37 — Bảng lượng tử hoá độ chói của JPEG, và cùng một khối ở hai mức chất lượng

- Tần số thấp chia cho số nhỏ: giữ gần nguyên; tần số cao chia cho số lớn: bị loại gần hết
- "Thanh chất lượng" của phần mềm chỉ là phép nhân cả bảng với một hệ số

Quét zig-zag

- Sau lượng tử hoá, khối chứa rất nhiều số 0 nhưng nằm rải rác nếu đọc theo hàng
- Quét zig-zag sắp 64 hệ số thành một dãy từ tần số thấp đến tần số cao
- Các số 0 dồn hết về cuối dãy, thuận lợi cho mã hoá entropy



(a) Đường quét zig-zag

7	26	−19	−11	−10	0	−6	0	−7	−6	0	−3	0	2	0	2
0	1	0	0	2	1	1	0	1	0	0	0	−1	0	0	0
0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ô nền xanh nhạt: 28 số 0 liên tiếp cuối dãy, thay bằng một ký hiệu "hết"

(b) Dãy 64 hệ số của khối ở Hình 8.36, theo thứ tự quét

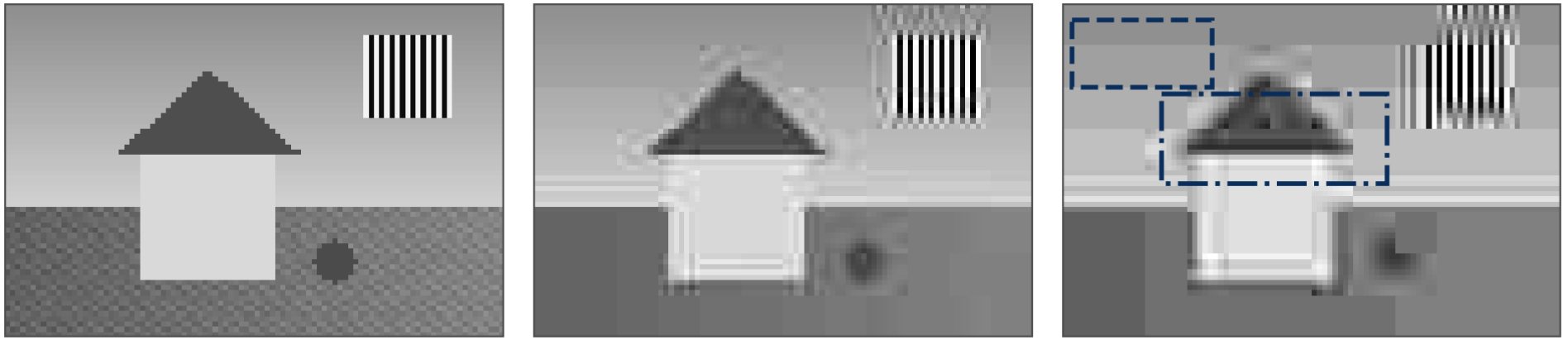
Hình 8.38 — Đường quét zig-zag và dãy hệ số thật của khối ở hình trước

Ba lớp mã hoá entropy (Bảng 8.32)

Kỹ thuật	Nguyên lý	Ví dụ
Mã chạy dài (RLC)	Thay chuỗi giá trị giống nhau bằng cặp (giá trị, số lần lặp)	Hai mươi tám số 0 cuối dãy chỉ cần một ký hiệu "hết"
Mã độ dài thay đổi (VLC)	Giá trị hay gặp gán mã NGẮN, giá trị hiếm gán mã DÀI	Huffman trong JPEG và MPEG-2; CAVLC trong H.264
Mã số học	Mã hoá cả dãy thành một số thực duy nhất, gần giới hạn entropy	CABAC trong H.264 và HEVC; hơn Huffman khoảng 10 – 15 %

- Cả ba đều không tổn hao: khai thác loại dư thừa thứ ba (mục 8.14) và khôi phục chính xác từng hệ số

Nhận ra lỗi nén bằng mắt



Hình 8.39 — Hai hình dạng đặc trưng của lỗi nén trong khung

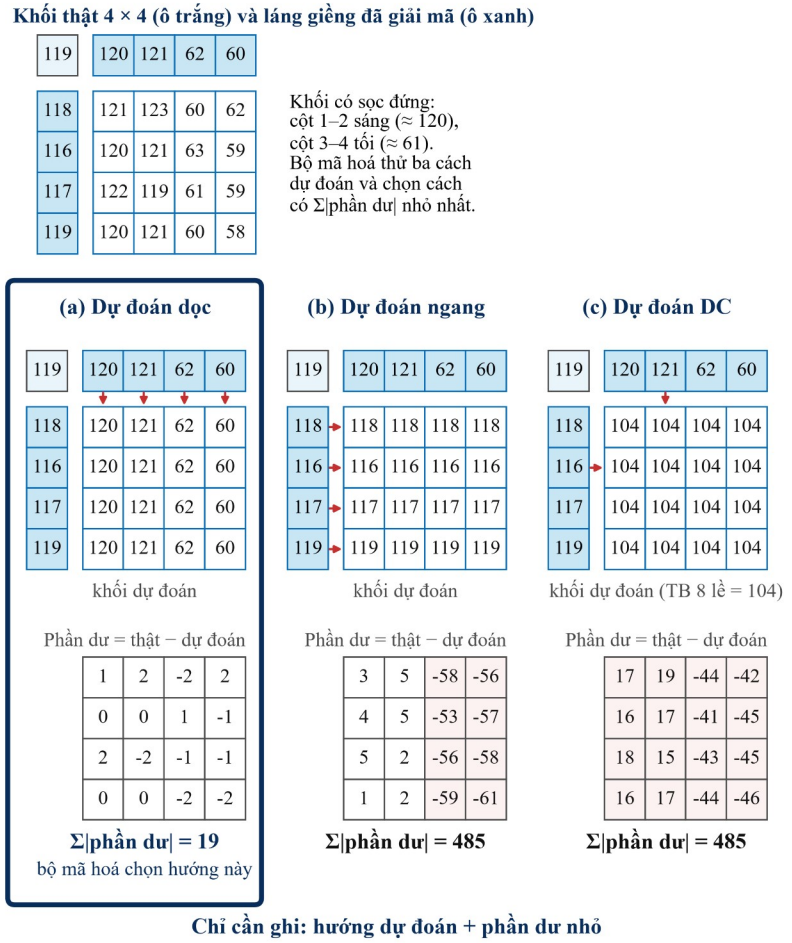
- Nhận ra các lỗi này cho biết tốc độ bit đang đặt quá thấp

Ba lỗi nén trong khung (Bảng 8.33)

Lỗi	Hình dạng	Nguyên nhân
Vỡ khối	Ô vuông 8 × 8 hiện rõ ở vùng lẽ ra chuyển sắc êm: bầu trời, phông vải, làn da	Mỗi khối lượng tử hoá độc lập nên mức trung bình các khối cạnh nhau lệch một bậc
Nhiễu muỗi	Chấm lấm tấm quanh mép sắc nét, nhất là chữ và cành lá trên nền trời	Mép sắc cần nhiều hệ số tần số cao; xoá bớt chúng sinh dao động dư quanh mép
Bệt chi tiết	Vùng hoa văn mịn — cỏ, tóc, vải — thành mảng phẳng	Toàn bộ hệ số tần số cao của khối bị xoá về 0

Dự đoán trong khung ở codec hiện đại

- Từ H.264, khối được dự đoán từ điểm ảnh lân cận đã giải mã (hàng trên, cột trái); chỉ phần dư được biến đổi
- Ví dụ khối 4 × 4 sọc đứng: dự đoán dọc cho tổng phần dư 19; ngang 485; DC 485
- Chọn hướng có tổng nhỏ nhất; luồng bit ghi hướng đã chọn và phần dư
- Lân cận dùng phải là điểm ảnh đã TÁI TẠO, vì bộ giải mã chỉ có bản tái tạo
- Số hướng tăng theo thế hệ: 9 (H.264), 35 (HEVC), 67 (VVC)



Hình 8.40 — Dự đoán trong khung trên khối 4 × 4

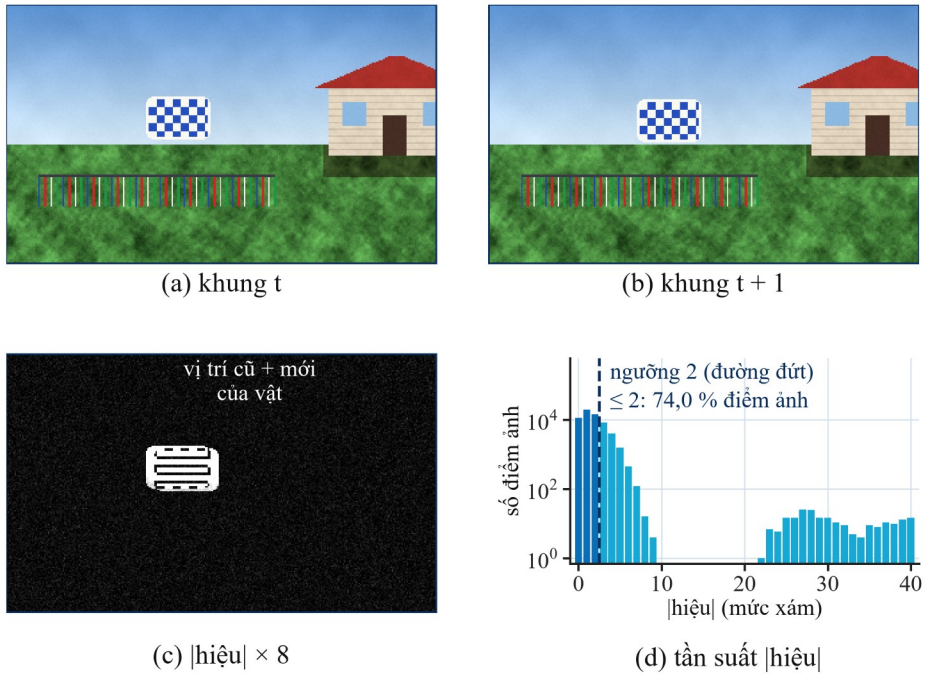
Những điều cần nhớ

- DCT chỉ viết lại 64 số; mất mát chỉ ở lượng tử hoá
- Bảng lượng tử hoá không đều: giữ tần số thấp, bỏ tần số cao; thanh chất lượng là phép nhân
- Zig-zag dồn số 0 về cuối; RLC, VLC, mã số học đều không tổn hao
- Ba lỗi: vỡ khối, nhiễu muỗi, bệt chi tiết — dấu hiệu tốc độ bit quá thấp
- Dự đoán trong khung (từ H.264) làm phần dư nhỏ hơn trước khi biến đổi

Nén liên khung

- Khai thác dư thừa thời gian: các khung liền nhau gần như giống nhau
- Ba cấp: lấy hiệu, dò chuyển động theo khối, phần dư qua đường ống nén ảnh
- Khung I, P, B, nhóm ảnh, độ trễ và tổng hợp số bit qua từng tầng

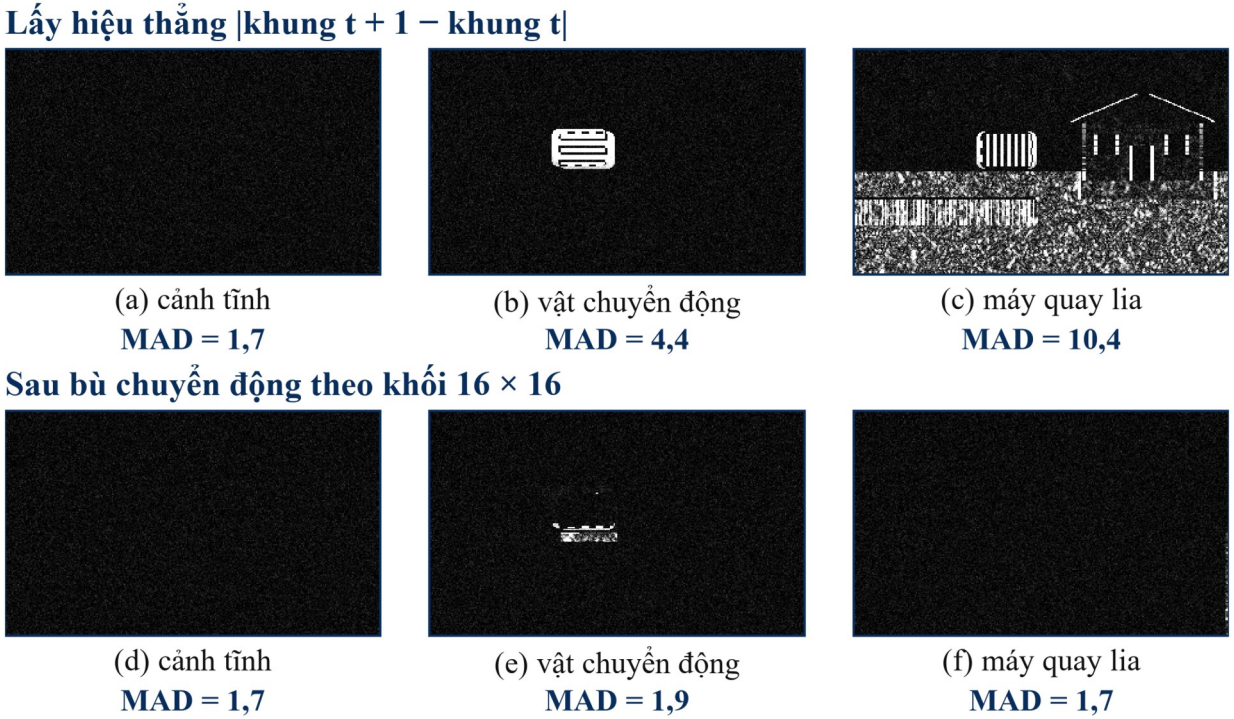
Hai khung liên tiếp giống nhau tới mức nào



Hình 8.41 — Hai khung liên tiếp của một cảnh mô phỏng: ảnh hiệu gần như đen

- 74,0 % điểm ảnh gần như không đổi (|hiệu| ≤ 2 trên thang 255)
- Entropy của khung: 7,46 bit/điểm; của ảnh hiệu: chỉ 3,37 bit/điểm — cùng hiện tượng ở mục 8.14.4, theo chiều thời gian

Cấp một: lấy hiệu thẳng giữa hai khung



Hình 8.42 — Lấy hiệu thẳng: tốt với cảnh tĩnh, hỏng khi có chuyển động; bù chuyển động đưa phần dư về mức nhiễu

- Độ lệch tuyệt đối trung bình (MAD): cảnh tĩnh 1,7; vật chuyển động 4,4; máy quay lia 10,4
- Có chuyển động thì phần dư lớn ngang một ảnh mới: cần xác định vật đã dịch đến đâu

Cấp hai: dò chuyển động theo khối [31]

- **1. Chia khối** 16 × 16; một vectơ cho mỗi điểm ảnh sẽ tốn hơn lợi thu được
- **2. Tìm trong cửa sổ** ± R điểm ảnh quanh vị trí cũ ở khung tham chiếu
- **3. Đo độ giống bằng SAD** với từng vị trí ứng viên
- **4. Chọn vectơ** có SAD nhỏ nhất; khối ở đó là khối dự đoán, hiệu với khối hiện tại là phần dư

Tổng giá trị tuyệt đối của hiệu — SAD (8.29)

Thước đo độ giống giữa khối hiện tại và khối ứng viên.

$$\mathrm{SAD}(d_x, d_y) = \sum_{i,j} \left| B_t(i,j) - B_{t-1}(i + d_x,\ j + d_y) \right|$$

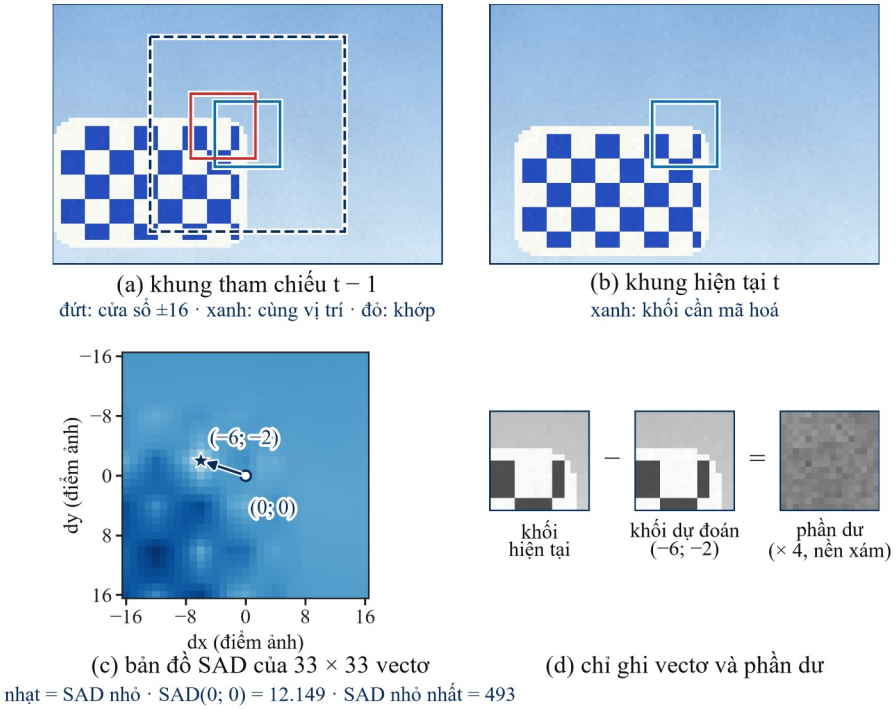
- B(t) là khối hiện tại, B(t−1) là khối ứng viên ở khung tham chiếu
- Vị trí SAD nhỏ nhất cho vectơ chuyển động (dx, dy)

Ví dụ khối 4 × 4 có cạnh đứng (Bảng 8.34)

Vị trí ứng viên (dx)	Một dòng khối tham chiếu	Hiệu tuyệt đối một dòng	SAD cả khối
0	10 · 10 · 10 · 80	0 · 0 · 70 · 0	280
+1	10 · 10 · 80 · 80	0 · 0 · 0 · 0	0
+2	10 · 80 · 80 · 80	0 · 70 · 0 · 0	280

- Dòng của khối hiện tại: 10, 10, 80, 80
- SAD nhỏ nhất bằng 0 ở dx = +1: vectơ (+1, 0), phần dư bằng 0 — cả khối chỉ cần một vectơ

Dò chuyển động trên khối 16 × 16 thật



Hình 8.43 — Cửa sổ tìm kiếm, bản đồ SAD với điểm cực tiểu, khối dự đoán và khối phần dư

- Cửa sổ ±16: SAD nhỏ nhất ở vectơ (−6, −2) bằng 493, so với 12 149 nếu không bù (0, 0)
- Vectơ trỏ từ khối hiện tại về khối giống nó ở khung tham chiếu nên ngược chiều với chiều dời của vật

Chi phí của tìm kiếm vét cạn

- Cửa sổ ±16: 33 × 33 = 1 089 vị trí; mỗi vị trí 256 phép hiệu ⇒ khoảng 279 nghìn phép mỗi khối
- Khung 1920 × 1080 có khoảng 8 100 khối ⇒ hơn hai tỉ phép mỗi khung
- Ở 25 hình/giây: khoảng 56 tỉ phép mỗi giây, chỉ cho dò chuyển động
- Bộ mã hoá thực tế dùng tìm kiếm nhanh: thu hẹp dần, bắt đầu từ vectơ khối láng giềng
- Đây là một nguồn gốc của chi phí tính toán ở mục 8.14.5

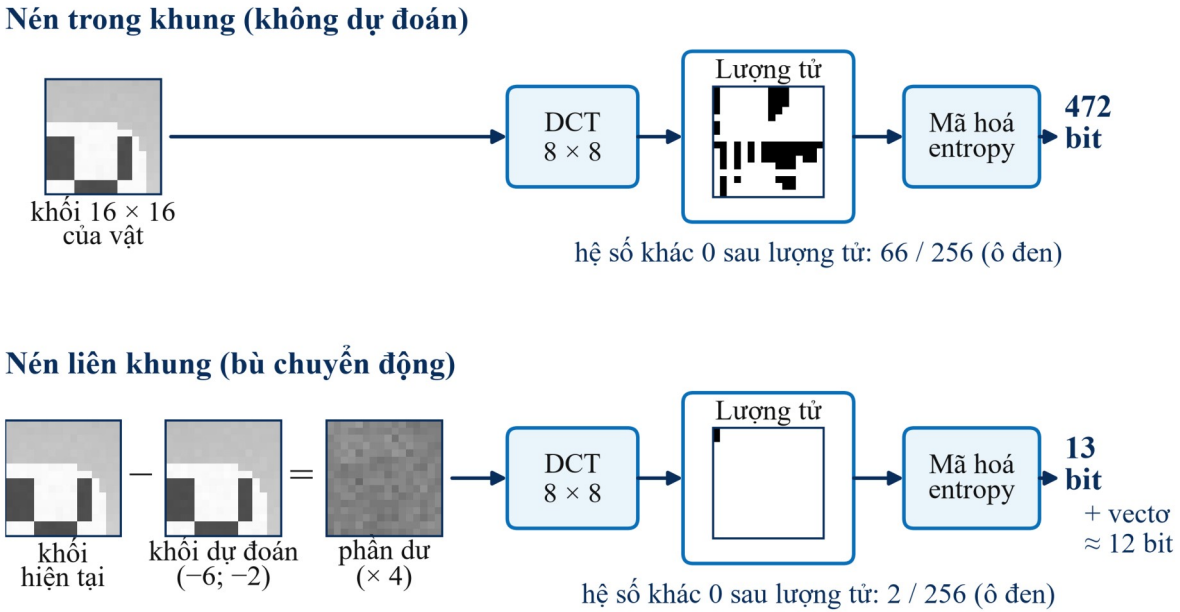
Cấp ba: phần dư qua đường ống nén ảnh (8.30)

Bộ giải mã dựng lại khối bằng cách cộng bản đoán với phần dư.

$$E = B - P \; ; \quad \hat{B} = P + \hat{E}$$

- P là khối dự đoán, B là khối thật, E là phần dư; Ê là phần dư đã qua đường ống và giải mã lại
- Phần dư đi qua đúng đường ống ở mục 8.16: DCT, lượng tử hoá, zig-zag, mã hoá entropy
- Vì nhỏ và phẳng nên sau lượng tử hoá hầu hết hệ số về 0

Phần dư qua cùng đường ống DCT

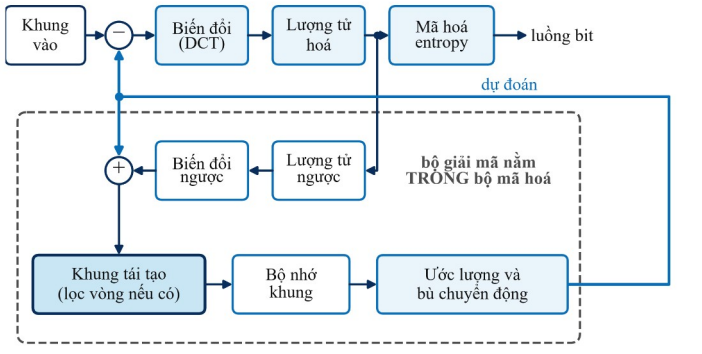


Tổng: 472 bit (trong khung) → 25 bit (liên khung) = 5 %

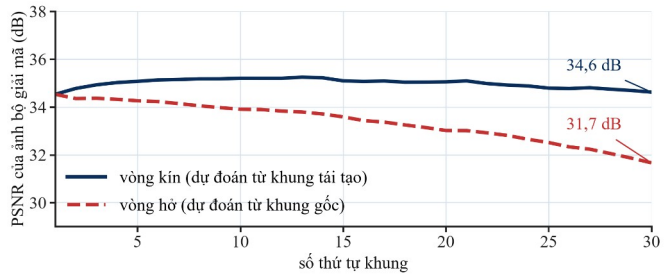
Hình 8.44 — Nén liên khung cần ít hệ số và ít bit hơn nén trong khung

- Cùng khối 16 × 16: trong khung 472 bit (66 hệ số khác 0); liên khung 25 bit (12 bit vectơ + 13 bit phần dư, 2 hệ số khác 0)
- Số bit tính trên bộ mã hoá thu nhỏ của hình: chỉ so sánh tương đối, không phải số bit của một chuẩn

Vòng phản hồi: bộ mã hoá chứa một bộ giải mã



(a) sơ đồ khối của bộ mã hoá lai



(b) mô phỏng 30 khung: lỗi cộng dồn nếu bộ mã hoá không giải mã

Hình 8.45 — Vòng phản hồi của bộ mã hoá lai (a) và hậu quả khi dự đoán từ khung gốc (b)

- Bộ giải mã chỉ có khung đã giải mã (đã lượng tử hoá), nên bộ mã hoá cũng phải dự đoán từ khung tái tạo
- Mô phỏng 30 khung: vòng hở giảm xuống 31,7 dB và xấu dần; vòng kín ổn định ở 34,6 dB

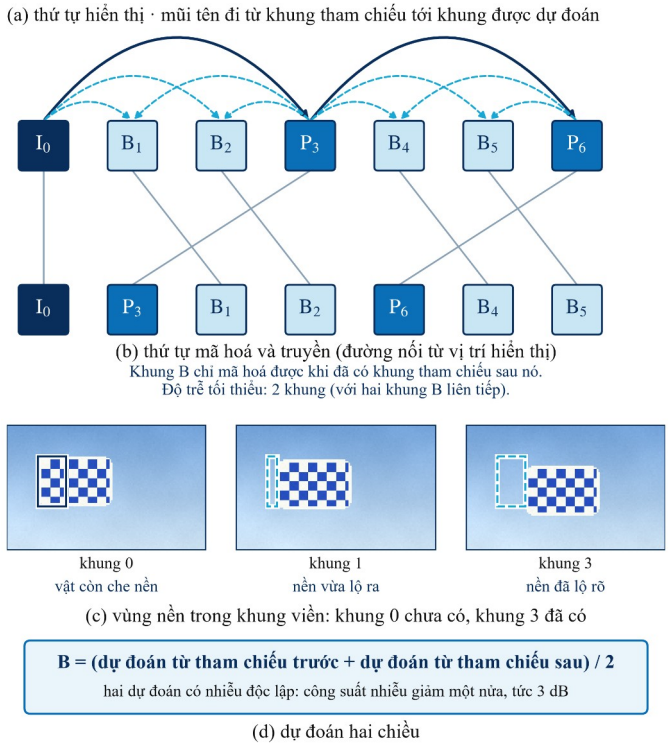
Hai hệ quả của vòng phản hồi

GHI NHỚ

Chuẩn chỉ quy định bộ giải mã; bộ giải mã phải chính xác từng bit.

- Chuẩn quy định cú pháp luồng bit và cách dựng lại hình, không quy định bộ mã hoá: bộ mã hoá được tự do cải tiến
- Vì vậy cùng chuẩn H.264 vẫn có bộ mã hoá tốt và bộ mã hoá kém
- Bộ giải mã của mọi nhà sản xuất phải ra cùng kết quả đến từng bit; sai một bit trong khung tham chiếu sẽ lan sang các khung sau

Khung B nhìn cả tương lai



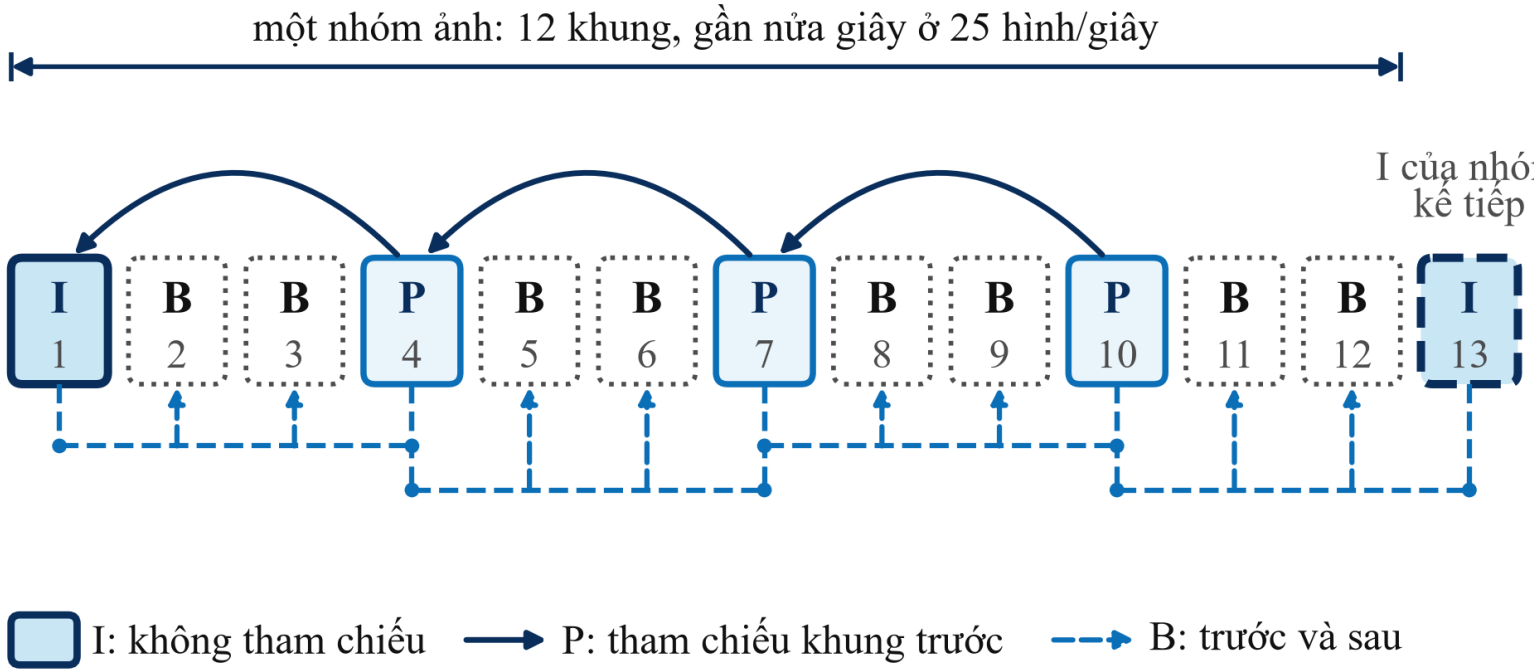
Hình 8.46 — Khung B tham chiếu cả quá khứ lẫn tương lai: thứ tự hiển thị khác thứ tự mã hoá

- Hiển thị I_0 B_1 B_2 P_3 được mã hoá theo thứ tự I_0 P_3 B_1 B_2; khung B gây độ trễ
- Trung bình hai bản đoán độc lập làm công suất nhiễu giảm một nửa, tốt hơn 3 dB

Ba loại khung (Bảng 8.35)

Loại	Tham chiếu	Kích thước tương đối	Đặc điểm
Khung I (Intra)	Không tham chiếu khung nào	100	Tự giải mã được; là điểm duy nhất có thể CẮT và TUA tới
Khung P (Predicted)	Khung đã giải mã TRƯỚC nó	≈ 34	Chỉ ghi phần khác so với khung tham chiếu
Khung B (Bi-directional)	CẢ khung trước lẫn khung sau	≈ 13	Nhỏ nhất, nhưng gây độ trễ vì phải chờ khung sau

Nhóm ảnh 12 khung



Hình 8.47 — Cấu trúc một nhóm ảnh 12 khung với quan hệ tham chiếu

- Nhóm ảnh: chu kỳ bắt đầu bằng khung I; hai tham số là độ dài nhóm và khoảng cách giữa hai khung tham chiếu liên tiếp

Chọn độ dài nhóm ảnh theo mục đích (Bảng 8.36)

Độ dài nhóm	Ưu điểm	Nhược điểm	Dùng khi
1 khung (chỉ I)	Cắt ở mọi khung; dựng rất nhẹ máy	Tệp lớn nhất	Codec dựng: ProRes, DNxHR, XAVC-I
12 – 15 khung	Cân bằng nén và khả năng cắt	Chỉ cắt được mỗi nửa giây	Phát sóng DVB-T2, đĩa DVD
48 – 250 khung	Nén hiệu quả nhất, tệp nhỏ nhất	Tua và dò kênh chậm; cắt phải mã hoá lại	Phát trực tuyến, lưu trữ nội dung hoàn thiện

Nhóm ảnh đóng và nhóm ảnh mở

Nhóm đóng

- Không khung nào tham chiếu ra ngoài nhóm
- Cắt và ghép nối ở ranh giới nhóm không cần mã hoá lại
- Chuẩn giao nộp của các đài thường bắt buộc nhóm đóng

Nhóm mở

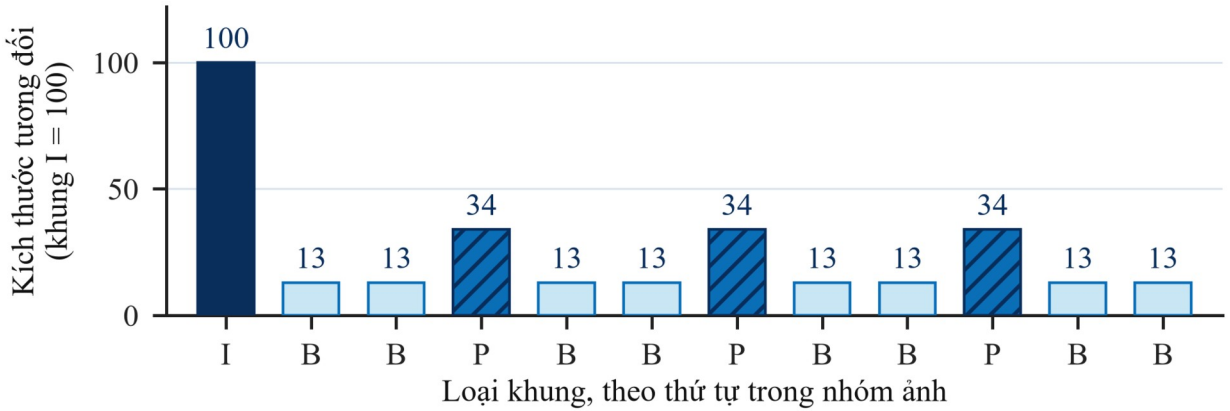
- Khung B đầu nhóm được tham chiếu khung I của nhóm trước
- Nén hiệu quả hơn đôi chút
- Không thể cắt ở ranh giới nhóm

Ba mức tinh vi của bù chuyển động (Bảng 8.37)

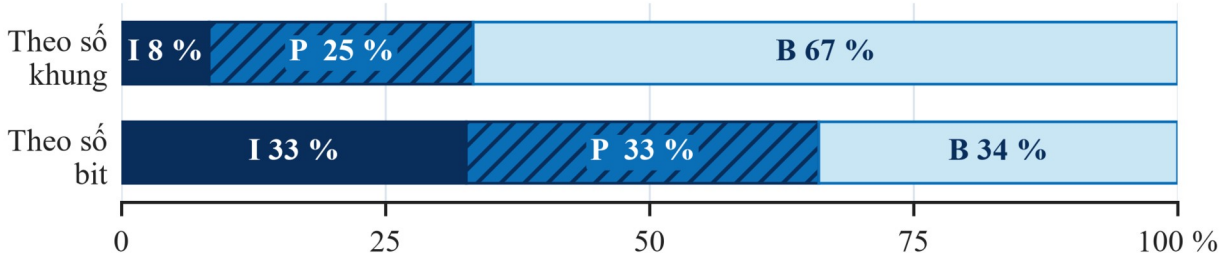
Kỹ thuật	Nội dung	Có trong chuẩn
Khối cố định, chính xác tới điểm ảnh	Khối 16 × 16, vectơ là số nguyên	MPEG-1, MPEG-2
Khối thay đổi, chính xác tới 1/4 điểm ảnh	Khối 16 × 16 xuống 4 × 4; vectơ có phần lẻ, nội suy giữa hai điểm ảnh	H.264 trở đi
Nhiều ảnh tham chiếu, khối tuỳ ý	Mỗi khối chọn tham chiếu từ nhiều khung, kể cả khung xa; hình dạng không nhất thiết vuông	HEVC, AV1, VVC

Bit phân bố ra sao giữa các loại khung

- Phải nêu rõ đang đếm theo số khung hay theo số bit
- Theo số khung (nhóm I B B P B B P B B P B B): I chiếm 8 %, P 25 %, B 67 %
- Theo số bit: khung I chỉ 8 % số khung nhưng chiếm khoảng MỘT PHẦN BA tổng số bit
- Khung I nặng gần gấp ba khung P và gần tám lần khung B
- Vì vậy nhóm ảnh dài giúp giảm tốc độ bit



(a) bit trung bình của mỗi khung



(b) tỉ lệ I, P, B theo số khung và theo số bit

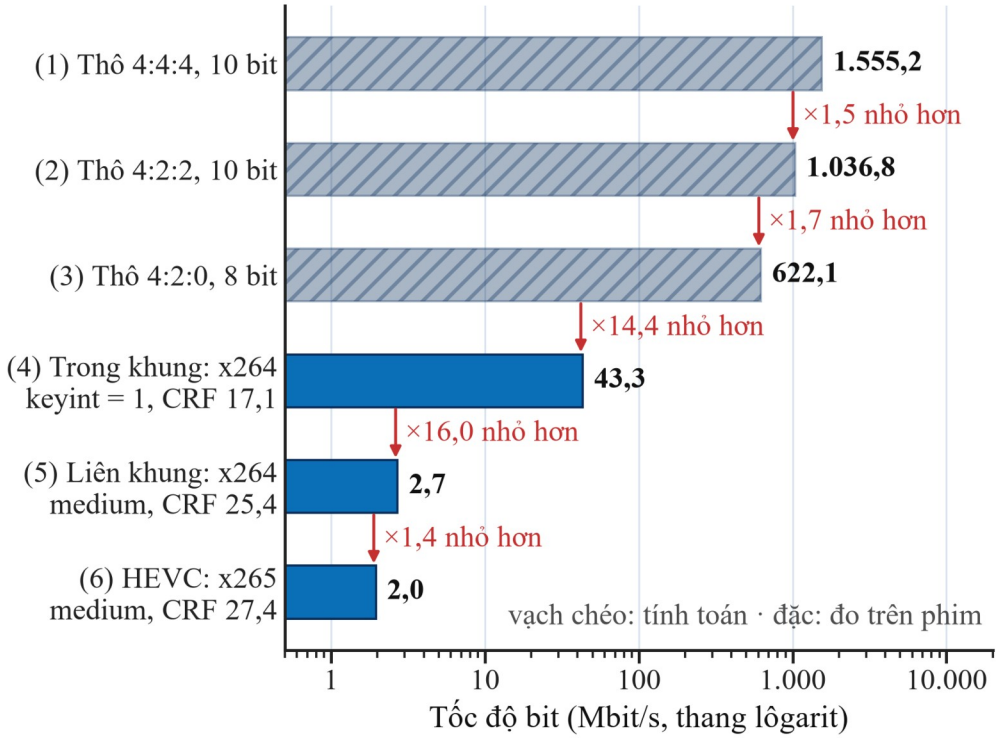
Hình 8.48 — Kích thước từng khung trong một nhóm ảnh

Đánh đổi giữa hiệu quả nén và độ trễ (Bảng 8.38)

Cấu hình	Độ trễ điển hình	Dùng cho
Chỉ khung I	Gần bằng không	Sản xuất trong trường quay, dựng
I và P, không có B	Một khung	Truyền hình trực tiếp, hội nghị truyền hình, truyền hình lưu động
I, P, B; nhóm 12 khung	Vài khung	Phát sóng thông thường
Nhóm dài, nhiều tầng khung B	Hàng chục khung	Phát trực tuyến nội dung hoàn thiện, lưu trữ

- Khung B phải chờ khung sau: độ trễ không chấp nhận được trong truyền hình trực tiếp

Cộng dồn: mỗi tầng giữ lại bao nhiêu bit



Hình 8.49 — Số bit giảm dần qua từng tầng kỹ thuật, đo trên đoạn phim thật ở cùng PSNR (trục lôgarit)

- 1080p25: 4:4:4 10 bit 1 555,2 → 4:2:2 1 036,8 → 4:2:0 8 bit 622,1 Mbit/s; trong khung 43,3; x264 liên khung 2,7; x265 2,0 Mbit/s
- Từng tầng: trong khung ≈ 14,4 lần; liên khung thêm ≈ 16,0 lần; HEVC thêm ≈ 1,4 lần so với H.264

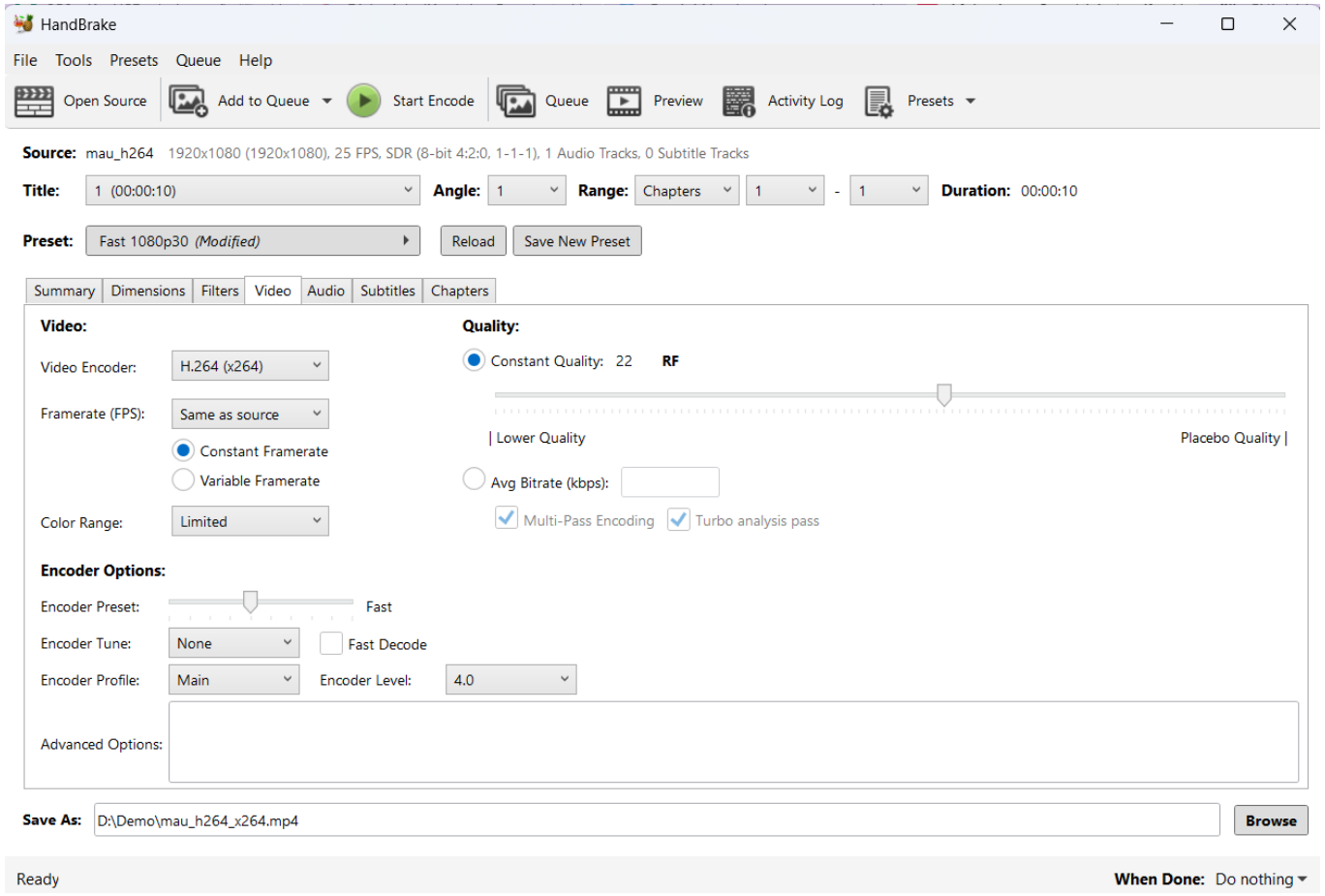
Những điều cần nhớ

- Hai khung liền nhau giống nhau tới mức entropy ảnh hiệu chỉ bằng một nửa khung
- Ba cấp: lấy hiệu, dò chuyển động theo khối (SAD), phần dư qua đường ống nén ảnh
- Bộ mã hoá chứa bộ giải mã (vòng phản hồi); chuẩn chỉ quy định bộ giải mã
- Khung I – P – B: kích thước tương đối 100 – 34 – 13; B nhỏ nhất nhưng gây độ trễ
- Nhóm đóng cho phép cắt; nhóm dài nén tốt hơn nhưng tua nặng
- Mỗi tầng kỹ thuật ứng với một loại dư thừa: cảm nhận, không gian, thời gian

Điều khiển tốc độ bit

- Ba chế độ: tốc độ bit cố định, thay đổi, chất lượng cố định
- Bộ đệm dung hoà giữa tốc độ mã hoá và tốc độ đường truyền
- Mã hoá một lượt và hai lượt; chọn và kiểm tốc độ bit

Chế độ chất lượng không đổi trong HandBrake



Hình 8.50 — Điều khiển theo chất lượng không đổi (RF) và các tuỳ chọn x264

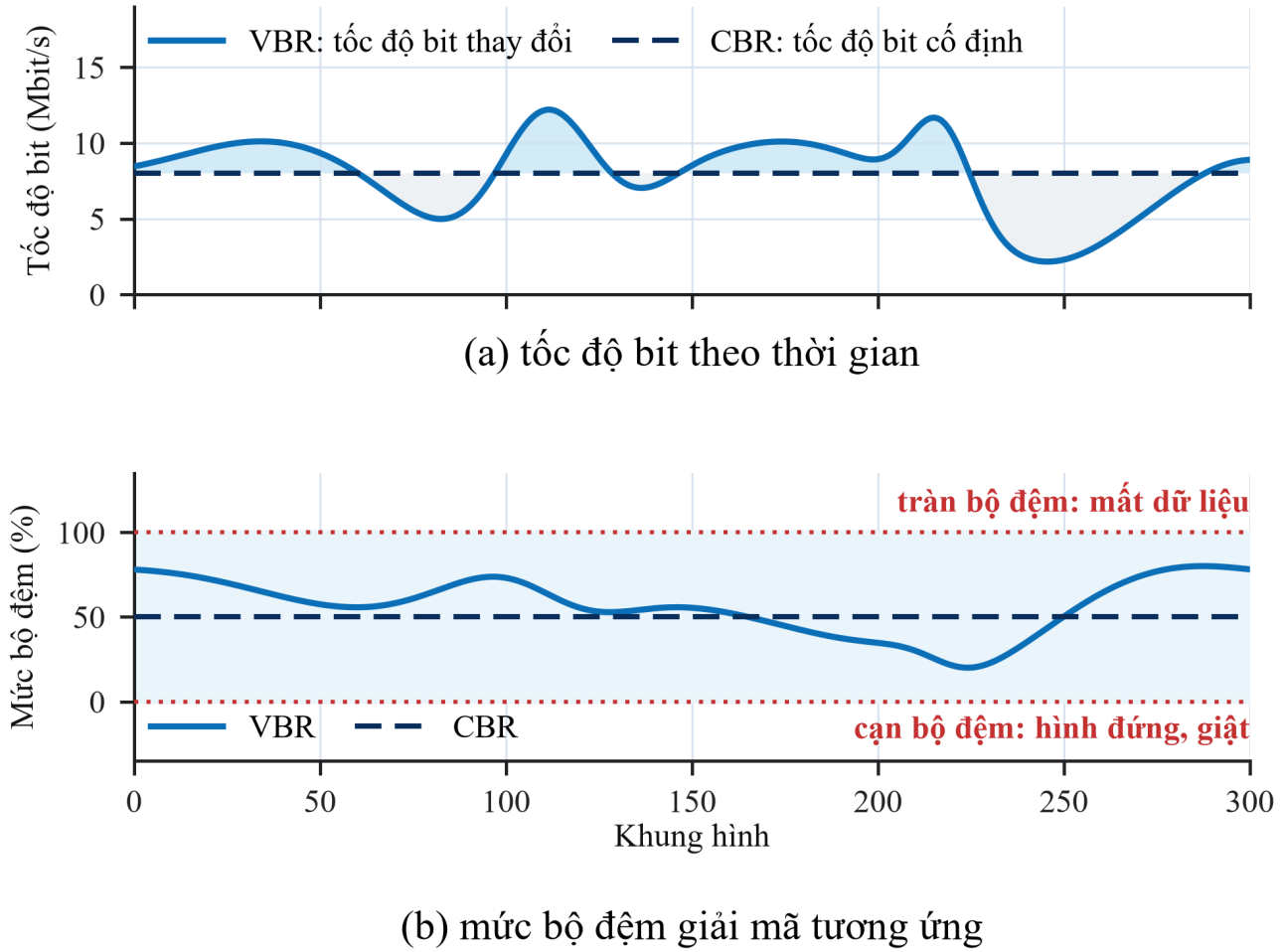
Ảnh chụp màn hình do người soạn thực hiện (HandBrake 1.11), 05/11/2026

- Thẻ Video: bộ mã hoá x264, mức tinh chỉnh (Encoder Preset), profile và level
- Chất lượng không đổi (RF): đặt mức chất lượng, tốc độ bit tự do

Ba chế độ điều khiển tốc độ bit (Bảng 8.39)

Chế độ	Cách làm	Ưu	Nhược
Tốc độ bit cố định (CBR)	Giữ tốc độ bit gần như không đổi	Dễ tính băng thông; bắt buộc với phát sóng và ghép kênh	Cảnh phức tạp thiếu bit nên vỡ hình; cảnh đơn giản thừa bit
Tốc độ bit thay đổi (VBR)	Cho tốc độ bit lên xuống theo độ phức tạp cảnh	Chất lượng đồng đều hơn ở cùng dung lượng trung bình	Không đoán trước đỉnh; khó ghép kênh
Chất lượng cố định	Đặt mức chất lượng, tốc độ bit tự do	Chất lượng ổn định nhất; hợp lưu trữ và bản trung gian	Không biết trước dung lượng tệp

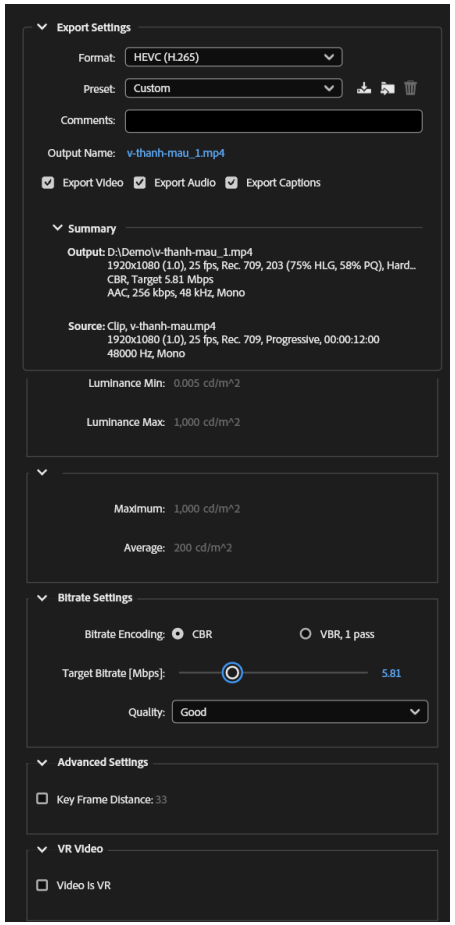
Hai chế độ và mức bộ đệm tương ứng



Hình 8.51 — Hai chế độ điều khiển tốc độ bit và mức bộ đệm tương ứng

Bộ đệm: chỗ dung hoà

- Bộ đệm cho phép tốc độ bit tức thời dao động quanh giá trị đặt, miễn không tràn và không cạn
- Ví dụ: Media Encoder, HEVC (H.265), CBR, tốc độ bit đích 5,81 Mbit/s
- Mục Bitrate Settings đặt tốc độ bit đích và hiện tóm tắt đầu ra



Hình 8.52 — Export Settings của Adobe Media Encoder (HEVC, CBR)

Ảnh chụp màn hình do người soạn thực hiện (Adobe Media Encoder 2026), 07/11/2026

Mức bộ đệm (8.31)

Mức bộ đệm tại thời điểm t.

$$B(t) = B(0) + \int_0^t [R_{\text{ch}} - R_{\text{enc}}(\tau)]d\tau \;; \quad 0 \le B(t) \le B_{\max}$$

- R(kênh) là tốc độ kênh, R(nén)(τ) là tốc độ bit tức thời của bộ mã hoá
- Mức bộ đệm là tích luỹ của chênh lệch giữa hai tốc độ
- Ràng buộc 0 ≤ B(t) ≤ Bmax: không cạn, không tràn

Hai sự cố của người xem, một nguyên nhân

CẢNH BÁO

Bộ đệm tràn làm vỡ hình; bộ đệm cạn làm hình đứng rồi nhảy.

- Tràn: bộ mã hoá đẩy vào nhanh hơn đường truyền chuyển đi; dữ liệu bị loại, người xem thấy hình vỡ hoặc mất khối
- Cạn: bộ giải mã lấy ra nhanh hơn tốc độ nhận; hình đứng lại rồi nhảy
- Tệp vượt kích thước bộ đệm quy định vẫn mở trên máy tính nhưng giật trên đầu thu; chỉ phát hiện bằng công cụ kiểm tra, không bằng cách xem thử

Mã hoá một lượt và hai lượt

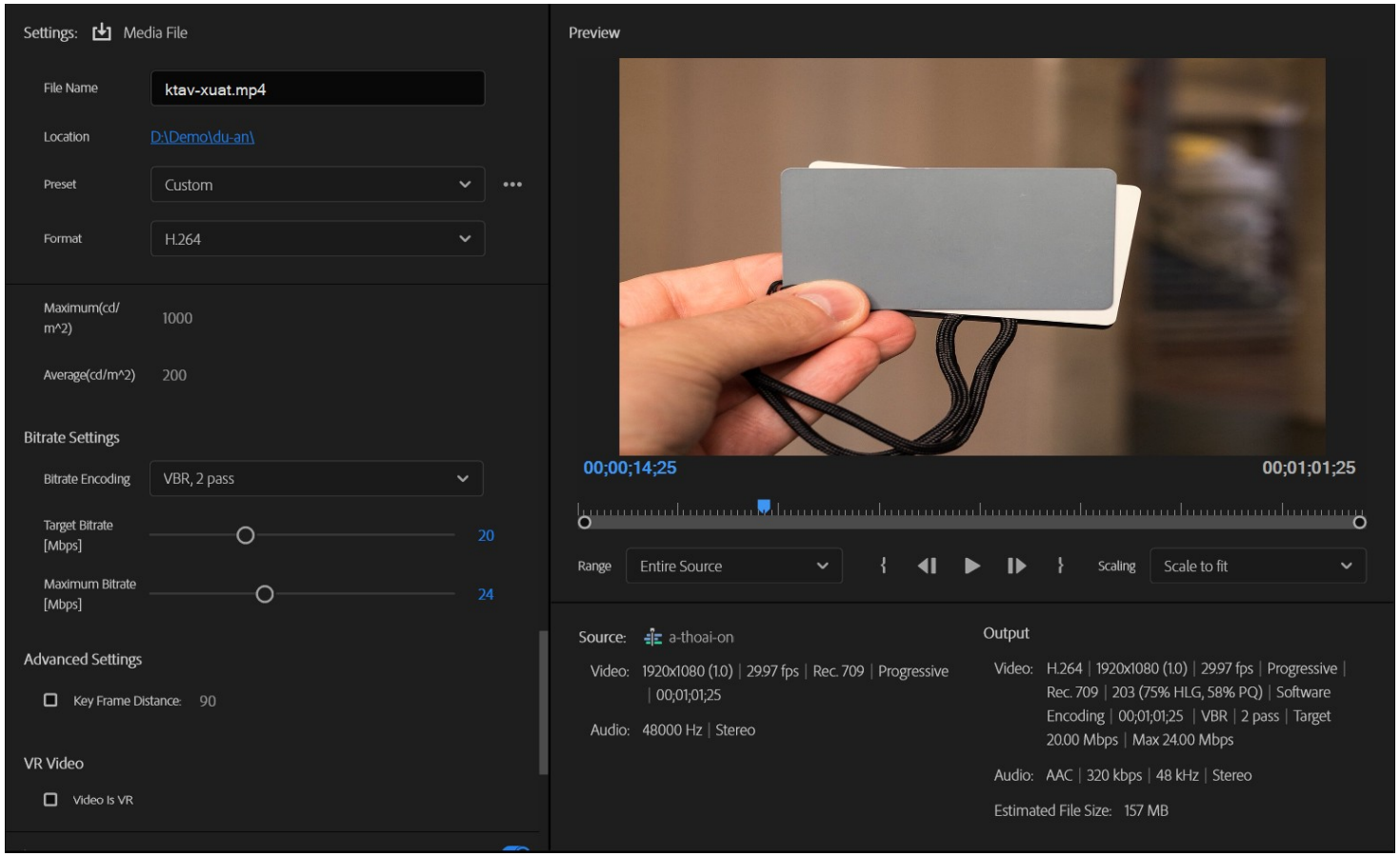
Một lượt

- Đi thẳng từ đầu đến cuối
- Phân bổ bit dựa trên những gì đã xử lý
- Bắt buộc khi phát trực tiếp

Hai lượt

- Lượt một phân tích toàn bộ tư liệu để biết đoạn khó và đoạn dễ
- Lượt hai mã hoá thật theo bản đồ đã có
- Tốt hơn rõ rệt ở cùng dung lượng; thời gian gấp đôi; chỉ với tệp hoàn chỉnh

VBR hai lượt trong Premiere Pro



Hình 8.53 — H.264 VBR hai lượt trong màn hình Export của Adobe Premiere Pro

Ảnh chụp màn hình do người soạn thực hiện (Adobe Premiere Pro 2026), 07/11/2026

- Bitrate Settings: VBR hai lượt
- Tốc độ bit mục tiêu 20 Mbit/s, tối đa 24 Mbit/s
- Có khoảng cách khung khoá và tóm tắt đầu ra

Tốc độ bit khởi điểm theo định dạng và codec (Bảng 8.40)

Định dạng	MPEG-2	H.264	HEVC	Ghi chú
SD 576i25	4 – 6 Mbit/s	2 – 3 Mbit/s	1 – 1,5 Mbit/s	Kênh SD trong bộ ghép kênh DVB-T2
HD 1080i25	15 – 18 Mbit/s	6 – 10 Mbit/s	3 – 5 Mbit/s	Kênh HD phát sóng
Full HD 1080p25	không dùng	8 – 12 Mbit/s	4 – 6 Mbit/s	Phát trực tuyến chất lượng cao
UHD 2160p50	không dùng	35 – 50 Mbit/s	15 – 25 Mbit/s	UHD phát trực tuyến

- Điểm khởi đầu cho nội dung chuyển động trung bình; không có con số dùng được cho mọi trường hợp
- Thể thao, ca nhạc nhấp nháy cần cao hơn; bản tin, phỏng vấn cần thấp hơn

Bốn bước kiểm tốc độ bit đã chọn (Bảng 8.41)

Bước	Việc làm	Điều cần thấy
1. Chọn đoạn KHÓ NHẤT	Đoạn nhiều chuyển động hoặc chi tiết mịn nhất: lia qua đám đông, mưa, khói, pháo hoa	Đoạn khó nhất đạt thì cả chương trình đạt
2. Mã hoá thử 30 giây	Mã hoá riêng đoạn ấy đúng cấu hình dự định	Không cần mã hoá cả chương trình để kiểm
3. Xem trên màn hình ĐÚNG CỠ	Màn hình khán giả sẽ dùng, khoảng cách thật	Màn hình máy tính đặt gần thấy lỗi nặng hơn; điện thoại thấy nhẹ hơn thực tế (mục 8.4)
4. Kiểm bằng số	Tính VMAF hoặc SSIM so với bản gốc; ghi làm mốc	Dần có bảng mốc riêng cho từng loại nội dung

Những điều cần nhớ

- CBR cho phát sóng; VBR cho chất lượng đồng đều; chất lượng cố định cho lưu trữ
- Bộ đệm dung hoà tốc độ mã hoá và tốc độ kênh: tràn gây vỡ hình, cạn gây đứng hình
- Hai lượt tốt hơn nhưng gấp đôi thời gian và chỉ với tệp hoàn chỉnh
- Bảng tốc độ bit chỉ là điểm khởi đầu; kiểm bằng đoạn khó nhất trên màn hình đúng cỡ

Bài thực hành số 2 — nhìn thấy nén

- Gồm bốn phần, dùng hai đoạn tư liệu tự quay
- Người học nhận ra lỗi nén bằng mắt, đo ảnh hưởng của nhóm ảnh, so sánh hai thế hệ codec và nhận biết tổn hao thế hệ
- Nộp bảng số liệu, ảnh chụp có chú thích và nhận xét ngắn

Mục tiêu và thiết bị

- Nhận ra vỡ khối, nhiễu muỗi bằng mắt; đo ảnh hưởng của độ dài nhóm ảnh tới việc tua và cắt
- So sánh hai thế hệ codec ở cùng tốc độ bit; nhận biết tổn hao thế hệ
- Máy tính có FFmpeg (dòng lệnh, miễn phí) hoặc phần mềm chuyển mã có giao diện; một phần mềm dựng để tua và cắt thử
- Hai đoạn tự quay, mỗi đoạn khoảng 30 giây: một đoạn tĩnh (phỏng vấn, chân máy) và một đoạn chuyển động mạnh (lia nhanh qua cây lá hoặc người chạy)

Phần 1 — Tìm ngưỡng vỡ hình

- Mã hoá mỗi đoạn bằng H.264 ở tốc độ bit giảm dần: 8, 4, 2, 1, 0,5 rồi 0,25 Mbit/s
- Xem từng bản trên màn hình lớn
- Ghi tốc độ bit mà tại đó mỗi đoạn bắt đầu lộ lỗi
- Giải thích vì sao đoạn chuyển động mạnh vỡ sớm hơn đoạn tĩnh rất nhiều

Phần 2 — Gọi tên lỗi

- Ở bản tốc độ bit thấp nhất, phóng to và chụp màn hình ba vùng
- Vùng bầu trời hoặc phông trơn; vùng có chữ hoặc cành lá trên nền sáng; vùng hoa văn mịn
- Chỉ ra và gọi đúng tên: vỡ khối, nhiễu muỗi, bệt chi tiết
- Giải thích nguyên nhân từng lỗi theo mục 8.16.5

Phần 3 — So sánh hai thế hệ codec

- Mã hoá cùng một đoạn bằng H.264 và HEVC ở CÙNG tốc độ bit, chọn mức đủ thấp để thấy khác biệt
- So sánh trực quan; nếu công cụ cho phép, tính thêm PSNR của cả hai
- Kết quả cần rút ra: HEVC cho chất lượng cao hơn rõ rệt ở cùng tốc độ bit
- Đo thử thời gian mã hoá của hai bản và nhận xét về cái giá phải trả

Phần 4 — Nhóm ảnh và tổn hao thế hệ

- Mã hoá cùng một đoạn với nhóm ảnh 1, 12 và 250 khung; so dung lượng ba tệp
- Thử tua và cắt trong phần mềm dựng
- Lấy bản 2 Mbit/s mã hoá LẠI năm lần liên tiếp ở cùng tốc độ bit
- Rút ra: nhóm dài cho tệp nhỏ nhất nhưng tua nặng, cắt kém linh hoạt; mã hoá lại năm lần làm chất lượng giảm rõ dù tốc độ bit không đổi

Yêu cầu nộp

- Bảng tốc độ bit ngưỡng của hai đoạn tư liệu
- Ba ảnh chụp màn hình có chú thích gọi tên lỗi
- Hai ảnh so sánh H.264 với HEVC kèm thời gian mã hoá
- Bảng so sánh ba độ dài nhóm ảnh
- Ảnh so sánh bản gốc với bản mã hoá lại năm lần
- Đoạn nhận xét ngắn về codec và tốc độ bit nên chọn cho loại nội dung nhóm thường thực hiện

Những điều cần nhớ

- Ngưỡng vỡ hình phụ thuộc mạnh vào mức chuyển động của nội dung
- Ba lỗi gọi đúng tên: vỡ khối, nhiễu muỗi, bệt chi tiết
- HEVC tốt hơn H.264 ở cùng tốc độ bit, đổi lại thời gian mã hoá lâu hơn
- Nhóm ảnh dài nhỏ tệp nhưng khó cắt; mã hoá lại nhiều lần gây tổn hao thế hệ

Kết thúc Phần C

- Tín hiệu chưa nén cần hơn 1 Gbit/s cho HD; nén khai thác bốn loại dư thừa, chỉ dư thừa cảm nhận gây mất mát
- Bốn việc: dự đoán, biến đổi, lượng tử hoá, mã hoá entropy; mất mát chỉ ở lượng tử hoá
- Nén liên khung: bù chuyển động theo khối, khung I – P – B, nhóm ảnh, vòng phản hồi
- Điều khiển tốc độ bit: CBR, VBR, chất lượng cố định; bộ đệm không tràn, không cạn
- Phần D: các họ chuẩn nén cụ thể — MPEG-2, H.264, HEVC, VVC, AV1 — được xây trên cùng bộ khung này