ĐẠI HỌC QUỐC GIA HÀ NỘI

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

ĐÀO THỊ HUỆ LÊ

CÁC GIẢI PHÁP NÂNG CAO HIỆU NĂNG MÃ HÓA VIDEO KHẢ CHUYỂN CHO HỆ THỐNG CAMERA GIÁM SÁT HÌNH ẢNH

LUẬN VĂN THẠC SĨ

NGÀNH CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ - VIỄN THÔNG

Chuyên ngành: Kỹ thuật viễn thông

HÀ NỘI, 2019

LỜI CAM ĐOAN

Tôi xin cam đoan đề tài "CÁC GIẢI PHÁP NÂNG CAO HIỆU NĂNG MÃ HÓA VIDEO KHẢ CHUYỂN CHO HỆ THỐNG CAMERA GIÁM SÁT HÌNH ẢNH" do TS. Hoàng Văn Xiêm và TS. Đinh Triều Dương hướng dẫn là công trình nghiên cứu của tôi. Các nội dung nghiên cứu, kết quả trong luận văn này là trung thực và không sao chép các công trình của người khác.

Tất cả các tài liệu tham khảo được sử dụng trong khóa luận này được ghi rõ nguồn gốc và ghi tên tác giả. Nếu có sai sót, tôi xin chịu hoàn toàn trách nhiệm.

Hà Nội, ngày tháng năm 2019

Tác giả

Đào Thị Huệ Lê

LỜI CẢM ƠN

Em xin bày tỏ lòng biết ơn sâu sắc đến hai thầy TS. Hoàng Văn Xiêm và TS. Đinh Triều Dương đã tận tình hướng dẫn và hỗ trợ em hoàn thành luận văn này.

Em cũng xin cảm ơn các thầy, cô giáo của khoa Điện tử - Viễn thông, Trường Đại học Công nghệ - Đại học Quốc gia Hà Nội đã giảng dạy cho em những kiến thức bổ ích trong suốt quá trình học tập tại trường.

Công trình này được tài trợ một phần từ đề tài KHCN cấp ĐHQGHN, Mã số đề tài: QG.19.22

Cuối cùng, con xin bày tỏ lòng biết ơn vô hạn với gia đình thân yêu của mình. Cảm ơn những người bạn thân thiết đã chăm sóc, khích lệ mình trong quá trình học tập nghiên cứu.

Trong quá trình thực hiện luận văn này, do kiến thức còn hạn chế, không thể tránh khỏi những thiếu sót, em rất mong nhận được những ý kiến góp ý của các thầy cô để em có thể tiếp tục hoàn thiện và phát triển đề tài này.

Hà Nội, tháng 5 năm 2019

Đào Thị Huệ Lê

MỤC LỤC

CÁC THUẬT NGỮ VÀ CHỮ VIẾT TẮT i

DANH MỤC HÌNH ẢNH/ HÌNH VẼ iv

DANH MỤC BẢNG BIỂU v

PHẦN MỞ ĐẦU 1

PHẦN NỘI DUNG 4

CHƯƠNG 1. GIỚI THIỆU CHUNG 4

1.1. Đặt vấn đề 4

1.2. Tổng quan về mã hóa video 4

1.2.1. Cấu trúc chung 6

1.2.2. Chức năng và vai trò một số module chính 6

CHƯƠNG 2. TỔNG QUAN VỀ MÃ HÓA VIDEO KHẢ CHUYỂN VÀ ĐẶC TÍNH CỦA VIDEO GIÁM SÁT 8

2.1. Mã hóa video khả chuyển thế hệ cũ 9

2.1.1. MPEG-4/FGS 9

2.1.2. SVC 9

2.2. Mã hóa video khả chuyển thế hệ mới 10

2.2.1. SHVC 10

2.2.2. DSVC 10

2.3. Phân tích đặc tính của video giám sát 11

CHƯƠNG 3. GIẢI PHÁP TRÊN SHVC 15

3.1. Giới thiệu 15

3.2. Cấu trúc SHVC 16

3.3. Phương pháp mã hóa liên quan 18

3.4. Đề xuất trên SHVC 19

3.4.1. Kiến trúc SSVC 19

3.4.2. Cấu trúc tham chiếu thời gian dài 21

3.4.3. Lựa chọn và cập nhật khung tham chiếu thời gian dài thích nghi 23

3.5. Kết quả giải pháp SSVC 24
CHƯƠNG 4. GIẢI PHÁP TRÊN DSVC **27**
4.1. Cấu trúc DVC 27
4.2. Cấu trúc DSVC 29
4.3. Đề xuất trên DSVC 32
4.3.2. Đề xuất tạo khung SI 35
4.4. Kết quả 36
4.4.1. Các điều kiện thử nghiệm 36
4.4.2. Đánh giá hiệu năng 37
4.4.3. Đánh giá độ phức tạp S-DSVC 38
4.4.3.1. Phân tích thành phần S-DSVC 39
4.4.3.2. S-DSVC so với SHVC 41
KẾT LUẬN **42**
TÀI LIỆU THAM KHẢO **43**

CÁC THUẬT NGỮ VÀ CHỮ VIẾT TẮT

A	
AVC: Advanced Video Coding (H.264/AVC)	Nén video tiên tiến
ALRS: Adaptive Long-term Reference Selection	Cơ chế cập nhật và lựa chọn tham chiếu dài hạn thích ứng
B	
BD Rate: Bjontegaard-Delta Rate	
BiME: Bi-directional Motion	Ước lượng chuyển động hai chiều
BL: Base Layer	Lớp cơ bản
C	
CM: Correlation model	Mô hình tương quan
D	
DCT: Discrete Cosine Transform	Biến đổi côsin rời rạc
DPB: Decoded Picture Buffer	Bộ đệm hình ảnh đã giải mã
DVC: Distributed Video Coding	Mã hóa video phân tán
DSVC: Distributed Scalable Video Coding	Mã hóa video khả chuyển phân tán
E	
EL: Enhancement Layer	Lớp nâng cao
F	
FD: Frame difference	Khác biệt giữa các khung
G	
GOP: Group of Pictures	Nhóm các khung hình

H	
HLS: High Level Syntax	Cú pháp cấp cao
I	
ILP: Inter Layer Prediction	Dự đoán liên lớp
L	
LD: Low-delay	Cấu trúc mã hóa độ trễ thấp
M	
MCTF: Motion Compensated Temporal Filtering	Lọc bù chuyển động thời gian
MV: Motion vector	Véc tơ chuyển động
MSE: Mean Squared Error	Sai số toàn phương trung bình
P	
PAD: pixel-averaged difference	Sự khác biệt trung bình điểm ảnh
PSNR: Peak Signal to Noise Ratio	Tỷ số tín hiệu cực đại trên nhiễu
Q	
QP: Quantization Parameters	Tham số lượng tử
R	
RD: Rate Distortion	
S	
S-DSVC: Surveillance – Distributed Scalable Video Coding	Mã hóa video phân tán khả chuyển cho giám sát
SAD: Sum of Absolute Difference	Tổng chênh lệch tuyệt đối
SHVC: Scalable High Efficiency Video Coding (H.265/SHVC)	Mã hóa video khả chuyển thế hệ mới
SI: Side Information	Thông tin phụ trợ

SNR: Signal to Noise Ratio	Tỷ số tín hiệu trên nhiễu
SVC: Scalable Video Coding (H.264/SVC)	Mã hóa video khả chuyển thế hệ cũ
SSVC: Surveillance Scalable Video Coding	Mã hóa video khả chuyển cho giám sát
U	
UHD: Ultra High Definition	Độ phân giải cực cao
V	
VSS: Visual Surveillance Systems	Hệ thống giám sát hình ảnh
W	
WZ: Wyner-Ziv	Khung WZ

DANH MỤC HÌNH ẢNH/ HÌNH VẼ

Hình 1.1. Mô hình chung của bộ mã hóa và giải mã hóa video (CODEC).....5
Hình 1.2. Minh họa sự tương quan về thời gian và không gian trong video.....5
Hình 1.3. Sơ đồ khối mô hình mã hóa video.....7

Hình 2.1. Lớp video mã hóa và các lớp giải mã ở các thiết bị khác nhau.....8
Hình 2.2. Hệ thống camera giám sát 11
Hình 2.3. PAD giữa các khung liên tiếp.....13
Hình 2.4. Ví dụ về các khung hình và trường vectơ chuyển động tương ứng trong video giám sát.....13

Hình 3.1. Cấu trúc SHVC 2 lớp.....17
Hình 3.2. Cấu trúc dự đoán theo low-delay P 19
Hình 3.3. Kiến trúc SSVC.....20
Hình 3.4. Cấu trúc dự đoán độ trễ thấp (LD) trong chuẩn mã hóa video thông thường (a) và trong SSVC (b) 22
Hình 3.5. Giải pháp ALRS 23
Hình 3.6. Khung đầu tiên của các video thử nghiệm 24
Hình 3.7. So sánh RD performance với trường hợp SSVC có cập nhật khung tham chiếu dài hạn.....26

Hình 4.1. Phương pháp nén video phân tán với thông tin phụ trợ 29
Hình 4.2. Cấu trúc DSVC.....30
Hình 4.3. Bộ mã hóa DSVC 30
Hình 4.4. Bộ giải mã DSVC.....31
Hình 4.5. Cấu trúc S-DSVC 33
Hình 4.6. Sơ đồ MCTF đề xuất 35
Hình 4.7. Khung đầu tiên của các video thử nghiệm 36
Hình 4.8. So sánh RD performance S-DSVC 38
Hình 4.9. So sánh thời gian mã hóa và giải mã.....39
Hình 4.10. Thời gian mã hóa – Các thành phần.....40
Hình 4.11. Thời gian giải mã – Các thành phần.....40
Hình 4.12. So sánh thời gian mã hóa.....41

DANH MỤC BẢNG BIỂU

Bảng 1. Tổng hợp các điều kiện kiểm tra SSVC.....25
Bảng 2. Tiết kiệm BD-Rate với SSVC.....26
Bảng 3. Điều kiện thử nghiệm S-DSVC 37
Bảng 4. Tiết kiệm BD-Rate S-DSVC.....37
Bảng 5. Cấu hình phần cứng để thử nghiệm S-DSVC.....39

PHẦN MỞ ĐẦU

1. Lý do chọn đề tài

Nén video là chủ đề đã và đang thu hút sự quan tâm của các nhà nghiên cứu trên thế giới. Ngày nay, các ứng dụng truyền thông đa phương tiện như truyền hình hội nghị, truyền hình quảng bá hay truyền hình Internet đã và đang đóng một vai trò không thể thiếu trong cuộc sống hiện đại. Có đến khoảng 87% dịch vụ quảng cáo hiện nay sử dụng các đoạn video clip và khoảng 85% lượng tắc nghẽn mạng vào năm 2019 sẽ nằm ở các luồng dữ liệu video [1].

Hiện nay, để đảm bảo an ninh ngày càng nhiều hệ thống giám sát (Visual Surveillance Systems - VSS) được lắp đặt. VSS là một hệ thống bao gồm các thiết bị điện tử được kết nối với nhau để ghi nhận hình ảnh tại địa điểm cần theo dõi và đưa hình ảnh tới người sử dụng bằng mạng Internet và hiển thị trên các thiết bị như tivi, máy tính hay điện thoại. Hệ thống giám sát hình ảnh được sử dụng ngày càng rộng rãi bởi chúng có thể hoạt động không biết mệt mỏi suốt ngày đêm. Chính vì sự gia tăng của các hệ thống giám sát và việc áp dụng rộng rãi trên toàn thế giới mà ta cần phải có giải pháp nén một cách hiệu quả lượng lớn video khổng lồ này để phục vụ việc truyền và lưu trữ.

Hiện tại có rất nhiều loại thiết bị hiển thị và các loại dịch vụ mạng với băng thông khác nhau mà các chuẩn mã hóa hiện tại như H.264 Advanced Video Coding (H.264/AVC) [4] hay H.265 High Efficiency Video Coding (H.265/HEVC) [2] không có khả năng mở rộng, các bitstream đầu ra chỉ có thể giải mã một lần, đồng thời chỉ cung cấp một mức chất lượng phù hợp cho các ứng dụng với một loại thiết bị nhất định. Do đó việc nghiên cứu các giải pháp mã hóa video hiệu quả và có khả năng mở rộng linh hoạt cho hệ thống camera giám sát hình ảnh là rất quan trọng.

Do vậy, luận văn này tập trung và nghiên cứu đề tài: **Scalable Video Coding Improvements for Visual Surveillance Systems (Các giải pháp nâng cao hiệu năng mã hóa video khả chuyển cho hệ thống camera giám sát hình ảnh)**. Các nội dung cụ thể trong đề tài sẽ được mô tả chi tiết ở các mục sau:

2. Mục đích nghiên cứu

Luận văn tập trung tìm hiểu và nghiên cứu về các giải pháp nâng cao hiệu năng mã hóa video khả chuyển cho hệ thống camera giám sát hình ảnh.

3. Nhiệm vụ nghiên cứu

- Tìm hiểu mô hình mã hóa video khả chuyển thế hệ mới H.265/SHVC
- Tìm hiểu mô hình mã hóa video khả chuyển phân tán DSVC
- Tìm hiểu phương pháp trích xuất thông tin từ lớp cơ bản áp dụng cho lớp nâng cao nhằm xây dựng thuật toán.
- Áp dụng thuật toán vào chương trình và kiểm thử đánh giá thuật toán.

4. Giới hạn phạm vi nghiên cứu

Luận văn tập trung nghiên cứu về mô hình mã hóa video khả chuyển phân tán DSVC, mô hình mã hóa video khả chuyển thế hệ mới H.265/SHVC và các giải pháp nâng cao hiệu năng mã hóa video cho hệ thống giám sát hình ảnh dựa trên các chuẩn mã hóa video này.

5. Đóng góp của luận văn

Luận văn nghiên cứu các giải pháp nâng cao hiệu năng mã hóa video khả chuyển cho hệ thống camera giám sát hình ảnh qua đó áp dụng kiểm thử so với chuẩn mã hóa HEVC và SHVC. Kết quả nghiên cứu trong khuôn khổ luận văn đã được công bố trên 04 bài báo tạp chí và hội nghị uy tín, bao gồm:

Journal(s):

1. **Le Dao Thi Hue**, Luong. P.V, Duong. D. T, and Xiem HoangVan, "Efficient and Low Complexity Surveillance Video Compression using Distributed Scalable Video Coding," VNU Journal of Science: Computer Science and Communication Engineering, Vol. 34, No. 1, pp. 38-51, 2018.

Conference(s):

1. Quang Hoang Van, Xiem HoangVan, **Le Dao Thi Hue**, Vien Dinh Du, and Vu Nguyen Hong, "Complexity Controlled Side Information Creation for Distributed Scalable Video Coding," Sigtelcom, Mar. 2019.

2. **Le Dao Thi Hue**, Giap PhamVan, Xiem HoangVan, "Adaptive Long-term Reference Selection for Efficient Scalable Surveillance Video Coding," IEEE 12th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC-2018), Sep. 2018.

3. **Le Dao Thi Hue**, D.T. Duong, and Xiem HoangVan, "HEVC based distributed scalable video coding for surveillance visual system," Nafosted Conference on Information and Computer Science, Hanoi, Oct. 2017.

6. Cấu trúc luận văn

Nội dung chính của luận văn được trình bày qua 4 chương:

Chương 1: Giới thiệu chung

Chương này trình bày về vấn đề nghiên cứu và tổng quan về mã hóa video.

Chương 2: Tổng quan về mã hóa video khả chuyển và đặc tính của video giám sát

Chương này trình bày về một số mô hình mã hóa video khả chuyển và phân tích đặc tính của video giám sát

Chương 3: Giải pháp trên chuẩn mã hóa video SHVC

Chương này trình bày về giải pháp nâng cao hiệu năng nén cho hệ thống giám sát hình ảnh cho chuẩn mã hóa video liên lớp thế hệ mới H.265/SHVC. Kết quả nghiên cứu của chương này được công bố tại:

[1] **Le Dao Thi Hue**, Giap PhamVan, Xiem HoangVan, "Adaptive Long-term Reference Selection for Efficient Scalable Surveillance Video Coding," IEEE 12th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC-2018), Sep. 2018.

Chương 4: Giải pháp trên chuẩn mã hóa video DSVC

Chương này trình bày về giải pháp nâng cao hiệu năng nén cho hệ thống giám sát hình ảnh bằng cách kết hợp chuẩn mã hóa dự đoán và mã hóa phân tán. Kết quả nghiên cứu của chương này được công bố tại:

[1] **Le Dao Thi Hue**, Luong. P.V, Duong. D. T, and Xiem HoangVan, "Efficient and Low Complexity Surveillance Video Compression using Distributed Scalable Video Coding," VNU Journal of Science: Computer Science and Communication Engineering, Vol. 34, No. 1, pp. 38-51, 2018.

[2] Quang Hoang Van, Xiem HoangVan, **Le Dao Thi Hue**, Vien Dinh Du, and Vu Nguyen Hong, "Complexity Controlled Side Information Creation for Distributed Scalable Video Coding," Sigtelcom, Mar. 2019.

[3] **Le Dao Thi Hue**, D.T. Duong, and Xiem HoangVan, "HEVC based distributed scalable video coding for surveillance visual system," Nafosted Conference on Information and Computer Science, Hanoi, Oct. 2017.

PHẦN NỘI DUNG

CHƯƠNG 1. GIỚI THIỆU CHUNG

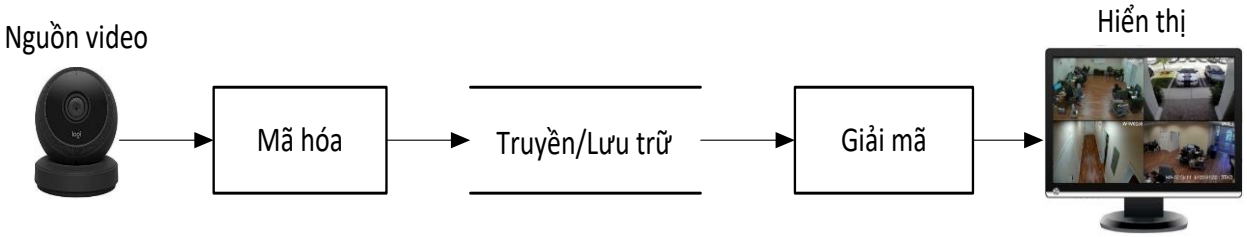
1.1. Đặt vấn đề

Cuộc cách mạng thông tin kỹ thuật số đã đem lại những thay đổi sâu sắc cho xã hội và đời sống con người. Trong cuộc sống hiện tại, không thể thiếu các thiết bị có khả năng chụp, xử lý, kết nối và truyền dữ liệu cho nhau thông qua Internet. Đặc biệt là hiện nay hệ thống giám sát hình ảnh để đảm bảo an ninh ngày càng trở nên phổ biến. Các hệ thống này có thể truyền tải thông tin hình ảnh và lưu trữ lâu dài. Bên cạnh những thuận lợi mà thông tin kỹ thuật số mang lại cũng tồn tại những thách thức và cơ hội cho quá trình đổi mới. Hiện nay các hình ảnh và video được chụp và tạo từ bất kì thiết bị nào đều cần lưu trữ để xem hoặc chia sẻ tới các thiết bị khác trong tương lai. Do đó, cần các giải pháp để lưu trữ và truyền video hiệu quả. Mục đích chính của các chuẩn mã hóa video là giúp nâng cao hiệu quả sử dụng băng tần và tăng khả năng lưu trữ thông tin video bằng cách làm giảm số bit cần để lưu trữ hay truyền đi.

Chương này sẽ trình bày tổng quan về mã hóa video cùng tình hình nghiên cứu về các chuẩn mã hóa video trên thế giới.

1.2. Tổng quan về mã hóa video

Một tập hợp hình ảnh với mỗi hình ảnh là một tập các điểm ảnh, mỗi điểm ảnh bao gồm thành phần độ sáng và thành phần màu sẽ tạo thành một video. Để chứa và truyền nhanh chóng lượng dữ liệu này, yêu cầu một lượng đáng kể bộ nhớ, do vậy việc nén các video này đóng một vai trò vô cùng quan trọng. Nén là quá trình mà dữ liệu biểu diễn thông tin được mã hóa với số lượng bit nhỏ hơn. Nén video là quá trình biểu diễn, kết hợp hay sắp xếp một tập hợp video số với số bit nhỏ hơn. Hình 1.1 biểu diễn quá trình từ thu thông tin đến hiển thị thông tin trong nén video, có hai khối chính là bộ mã hóa và bộ giải mã, chúng thường được gọi là CODEC. Bộ mã hóa biến đổi từ dữ liệu nguồn thành một dạng nén để lưu trữ hoặc truyền tải và bộ giải mã biến đổi dữ liệu đã nén thành video dữ liệu gốc.



Hình 1.1. Mô hình chung của bộ mã hóa và giải mã hóa video (CODEC)

Nén dữ liệu là việc loại bỏ dữ liệu dư thừa, ví dụ các thành phần không cần thiết để có thể tái hiện lại dữ liệu một cách đúng nhất. Rất nhiều loại dữ liệu có chứa dư thừa tĩnh và có thể nén một cách hiệu quả bằng cách nén không thất thoát dữ liệu, dữ liệu tái cấu trúc ở đầu ra của bộ giải mã là bản sao lưu hoàn hảo của dữ liệu gốc. Tuy nhiên, cách nén này chỉ mang lại hiệu quả nén ở mức trung bình.

Để đạt hiệu quả nén cao, hầu hết các phương thức mã hóa video đều khai thác dư thừa về mặt không gian và thời gian. Chuỗi hình ảnh tạo thành video được gọi là các khung hình (frames). Thường có sự tương quan giữa các khung hình với nhau về mặt thời gian khi chúng ta chụp lại trong cùng một khoảng thời gian. Các khung liền kề thường có độ tương quan cao đặc biệt khi tốc độ lấy mẫu (frame rate) cao. Về mặt không gian, thường có sự tương quan cao giữa các điểm ảnh gần nhau, tức là giá trị điểm ảnh tại những điểm gần nhau thường giống nhau (Hình 1.2).

Hình 1.2. Minh họa sự tương quan về thời gian và không gian trong video

Chuẩn mã hóa video phổ biến như MPEG-2/Video [3], H.264/AVC [4], H.265/HEVC [2] hay SHVC [6] đều được xây dựng trên nền tảng khai thác tính tương quan cao về mặt thời gian, không gian và thống kê của các khung hình video cũng như nội dung video. Các chuẩn mã hóa video đều sử dụng dự đoán bù chuyển động dựa trên khối, biến đổi cosin rời rạc, lượng tử hóa và mã hóa entropy. Phần tiếp theo sẽ trình bày các thành phần cơ bản của kiến trúc mã hóa video.

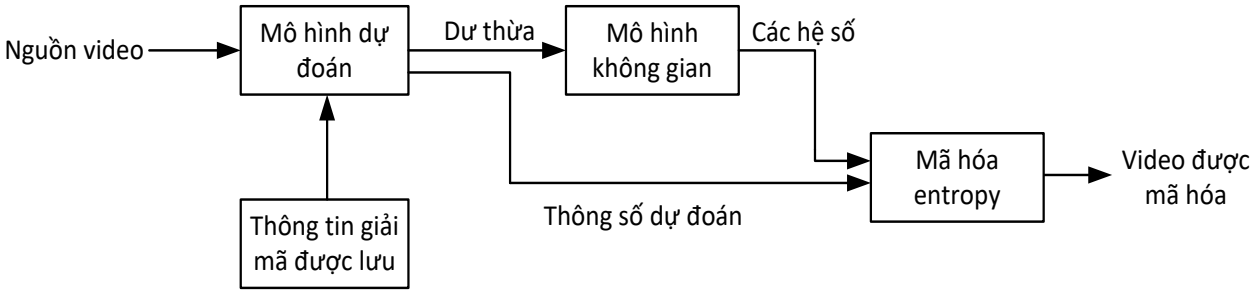
1.2.1. Cấu trúc chung

Một bộ mã hóa và giải mã video (còn gọi là CODEC) mã hóa chuỗi video hoặc ảnh nguồn thành dạng nén và giải mã nó thành video bản sao hoặc xấp xỉ so với chuỗi video/ảnh gốc. Quá trình mã hóa được gọi là mã hóa video không tổn thất nếu như video được giải mã giống hệt như video gốc, ngược lại, nếu như video giải mã có sự khác biệt so với video gốc, quá trình này được gọi là mã hóa video có tổn thất.

CODEC đại diện cho bản video gốc bằng một mô hình (một mô hình mã hóa hiệu quả là có thể tái cấu trúc thành video xấp xỉ bản gốc). Mô hình lý tưởng phải sử dụng ít dung lượng nhất và có độ tin cậy cao nhất có thể. Có hai mục tiêu nén là hiệu năng nén và chất lượng nén cao nhưng chúng thường xung khắc nhau vì một bộ nén hiệu năng cao sẽ tạo ra tốc độ bít (bitrate) thấp thường làm giảm chất lượng hình ảnh khi giải mã.

1.2.2. Chức năng và vai trò một số module chính

Mã hóa video (Hình 1.3) bao gồm 3 chức năng chính: mô hình dự đoán, mô hình không gian và mã hóa entropy. Đầu vào của mô hình dự đoán là chuỗi video chưa nén (video thô – “raw”). Mô hình dự đoán giả định giảm dư thừa về mặt thời gian bằng việc khai thác sự khác biệt về thông tin giữa các khung lân cận, thường bằng việc xây dựng dự đoán của khối (block) hoặc khung (frame) trong video hiện tại. Ở MPEG-4 và H.264/AVC, việc dự đoán được thực hiện từ một hay nhiều khung trước hoặc sau đó và được cải thiện bằng việc dự đoán bù chuyển động giữa các khung (motion compensated prediction). Đầu ra của mô hình dự đoán là các khung dư thừa (được tạo ra bằng việc trừ khung hiện tại gốc cho khung dự đoán) và một tập các tham số, ví dụ như một tập hợp các vectơ chuyển động mô tả việc bù chuyển động.



Hình 1.3. Sơ đồ khối mô hình mã hóa video

Các khung dư thừa là đầu vào của mô hình không gian, mô hình này khai thác sự tương quan giữa các mẫu lân cận và khung dư thừa để giảm sự dư thừa về mặt không gian bằng cách biến đổi dư thừa theo hệ số và lượng tử hóa kết quả. Hệ số được lượng tử hóa để loại bỏ các giá trị không quan trọng, giữ lại số lượng nhỏ các hệ số có giúp cho khung dư thừa gọn nhẹ hơn. Đầu ra của mô hình không gian là một tập các hệ số lượng tử biến đổi.

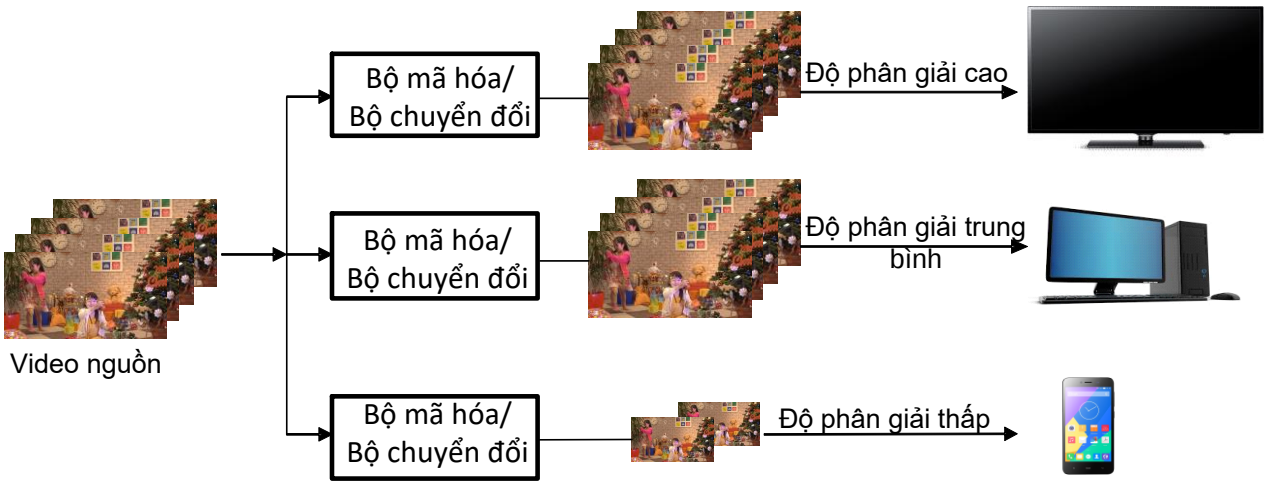
Các thông số của mô hình dự đoán (ví dụ chế độ dự đoán, các vectơ chuyển động) và mô hình không gian (các hệ số) được nén bởi mô hình mã hóa entropy. Mô hình này loại bỏ các dư thừa tĩnh trong dữ liệu (ví dụ các vectơ thường xuất hiện và các hệ số bằng mã hóa nhị phân ngắn hơn) và tạo ra một tập tin hay bitstream nén có thể truyền hoặc lưu trữ. Chuỗi được nén bao gồm các thông số dự đoán, các hệ số dư thừa và một số thông tin phụ trợ khác.

Bộ giải mã tái cấu trúc khung video từ bộ bitstream đã nén. Các hệ số và các thông số dự đoán được giải mã từ bộ giải mã entropy sau đó mô hình không gian được giải mã để tái tạo khung dư thừa. Bộ giải mã sử dụng các thông số dự đoán, cùng với các thông tin giải mã trước đó để tạo dự đoán cho khung hiện tại và khung tái tạo của chính nó bằng cách cộng khung dư thừa với khung dự đoán.

CHƯƠNG 2. TỔNG QUAN VỀ MÃ HÓA VIDEO KHẢ CHUYỂN VÀ ĐẶC TÍNH CỦA VIDEO GIÁM SÁT

Ngày nay, với sự phát triển của công nghệ video số và tiến bộ trong cơ sở hạ tầng truyền thông đã thúc đẩy sự tương tác của các ứng dụng truyền thông đa phương tiện như truyền hình hội nghị thời gian thực, website video trực tuyến hay TV, di động. Bên cạnh đó, người sử dụng luôn muốn chất lượng video tốt nhất ở bất cứ đâu và với mọi tình trạng của đường truyền mạng. Do đó, các đặc tính của máy thu như tốc độ bit, độ phân giải và tỷ lệ khung hình phải phù hợp với tốc độ truyền dẫn để cung cấp chất lượng video tốt nhất cho người sử dụng. Trên cùng một đường dẫn video thường chỉ cung cấp một mức chất lượng cho các thiết bị nhất định. Ngoài ra, luồng video trực tiếp còn phải đối mặt với hiện tượng mất tín hiệu ở mạng không dây. Vì vậy những mạng không đồng nhất này tạo ra một vấn đề không nhỏ cho các bộ mã hóa video truyền thống.

Để khắc phục nhược điểm này, mã hóa video khả chuyển được giới thiệu. Nguyên lý của bộ mã hóa video này là chia đoạn video truyền thống thành đa luồng, bao gồm thành phần riêng biệt và bổ sung, thường được gọi là các lớp.



Hình 2.1. Lớp video mã hóa và các lớp giải mã ở các thiết bị khác nhau

Hình 2.1 cho thấy video đầu vào được chia thành ba lớp. Người nhận có thể chọn và giải mã lớp khác nhau tương ứng với các đặc tính của mạng và thiết bị được sử dụng.

Trong các mô hình mã hóa video khả chuyển, thông tin video được mã hóa và giải mã theo nhiều lớp khác nhau bao gồm một lớp cơ bản (Base layer) và một hoặc nhiều lớp nâng cao (Enhancement layers). Có hai hướng tiếp cận chính trong việc xây dựng các mô hình mã hóa video khả chuyển: mô hình mã hóa video khả chuyển dự đoán (Predictive scalable video coding) và mô hình mã hóa video khả chuyển phân tán (Distributed scalable video coding).

2.1. Mã hóa video khả chuyển thế hệ cũ

Mã hóa video khả chuyển thế hệ cũ bao gồm các mô hình mã hóa video khả chuyển như chuẩn Fine Granularity Scalability (MPEG-4/FGS) [7] hay chuẩn Scalable Video Coding (SVC) [5] được mở rộng từ các chuẩn mã hóa video thế hệ cũ như chuẩn MPEG-2/Video [3], H.264/AVC [4].

2.1.1. MPEG-4/FGS

Chuẩn này ra đời năm 2001, các lớp nâng cao được thiết kế để mã hóa sự khác biệt giữa ảnh video gốc và ảnh giải mã từ các lớp thấp hơn, sử dụng phương pháp mã hóa theo từng lớp thông tin (bitplane) của các hệ số biến đổi côsin rời rạc (Discrete Cosine Transform - DCT). MPEG-4/FGS được thiết kế để phát trực tiếp video qua mạng truyền thông tốt hơn. Với chuẩn FGS, luồng video có thể được cắt ngắn một cách linh hoạt ở mức độ chi tiết rất tốt để thích ứng với các tài nguyên mạng có sẵn. Nhược điểm chính của chuẩn mã hóa MPEG-4/FGS là việc chưa khai thác được toàn bộ thông tin tại các lớp thấp hơn. Do vậy, chuẩn MPEG-4/FGS không mang lại hiệu năng mã hóa cao nên đã không được sử dụng phổ biến.

2.1.2. SVC

SVC thực chất là một phần mở rộng của H.264/AVC, ra đời năm 2007. Chuẩn mã hóa video SVC của luồng video chất lượng cao còn có thể chứa một hoặc nhiều luồng con. Luồng bit con có thể hiển thị ở độ phân giải thấp hơn (màn hình nhỏ hơn) hoặc với tốc độ khung thấp hơn. Bộ giải mã video SVC cho phép truyền video với nội dung không bị suy giảm giữa các đầu ra khác nhau, ví dụ giữa máy tính xách tay và thiết bị di động. Nhờ khả năng này, SVC có thể được ứng dụng trong các ứng dụng thực tế như truyền hình video

trên các thiết bị di động [8] hay phân tách dòng video thời gian thực giữa những người dùng ngang hàng [9].

2.2. Mã hóa video khả chuyển thế hệ mới

Mã hóa video khả chuyển thế hệ mới cũng giống như các mô hình mã hóa video khả chuyển thế hệ cũ, thông tin từ các video cũng được mã hóa và giải mã theo nhiều lớp khác nhau bao gồm một lớp cơ bản và một hoặc nhiều lớp nâng cao. Có hai hướng xây dựng mô hình là: *mô hình mã hóa video khả chuyển dự đoán* và *mô hình mã hóa video khả chuyển phân tán*. Tương ứng với hai hướng mô hình, phần này giới thiệu về chuẩn mã hóa video khả chuyển dự đoán, Scalable High Efficiency Video Coding (SHVC) [6] là phần mở rộng của chuẩn H.265/HEVC [2] và chuẩn mã hóa video khả chuyển phân tán, Distributed Scalable Video Coding (DSVC).

2.2.1. SHVC

Các nghiên cứu đã cho thấy H.265/HEVC đã giảm 52% lượng bit cần mã hóa ở độ phân giải video là 480p và có thể lên đến 64% ở độ phân giải 4K UHD khi so sánh với chuẩn mã hóa tiền nhiệm là H.264/AVC mặc dù H.265/HEVC mang lại chất lượng hình ảnh tốt hơn đáng kể khi nén với cùng kích thước video hoặc cùng tốc độ bit (Bitrate). Năm 2014, chuẩn mã hóa video khả chuyển dự đoán SHVC ra đời, được biết đến là phần mở rộng của chuẩn HEVC. SHVC hỗ trợ khả năng mở rộng theo thời gian, không gian và chất lượng video. Cũng tương tự như tiêu chuẩn SVC trước đây, SHVC cung cấp một cơ chế cho việc mã hóa video đa lớp, mỗi lớp đại diện cho một chất lượng hay độ phân giải nhất định cho cùng một bối cảnh của video. Các nghiên cứu gần đây hầu như chỉ tập trung vào việc phát triển mô hình mã hóa video khả chuyển dự đoán do hiệu năng mã hóa của chuẩn này có nhiều tiềm năng để nâng cao. Trong đó, hướng nghiên cứu chủ yếu là tạo ra các công cụ dự đoán tốt hơn nhằm sử dụng một cách có hiệu quả dữ liệu từ các lớp thấp hơn.

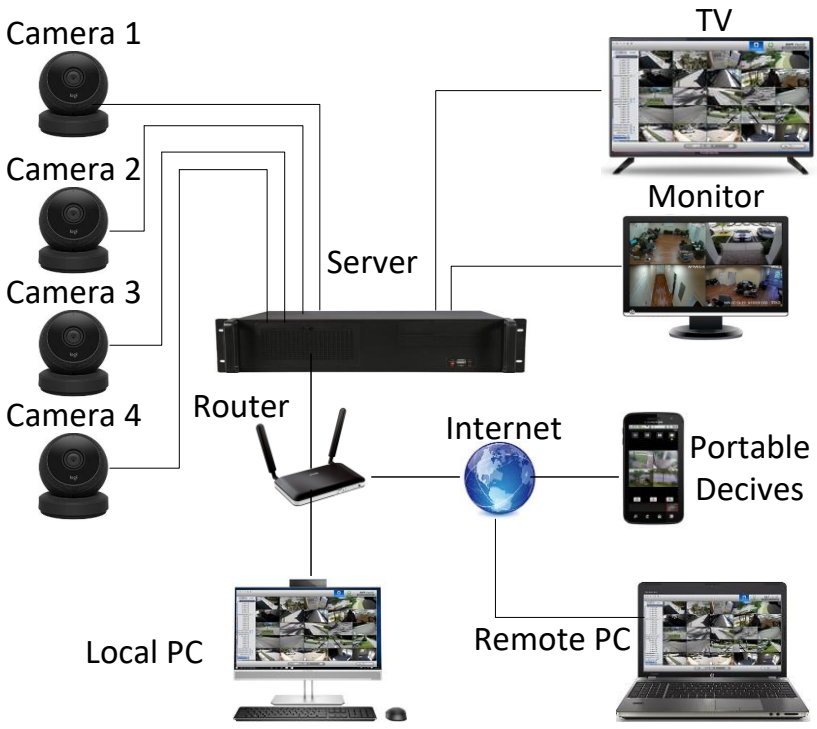
2.2.2. DSVC

Mô hình này được xây dựng dựa trên sự kết hợp hiệu quả các thế mạnh của hai kỹ thuật mã hóa video mới nhất hiện tại: H.265/HEVC (High Efficiency Video Coding) [2] và DVC (Distributed video coding - Mã hóa video phân tán) [10] nhằm tạo ra một mô hình mã hóa video mới, thống nhất và mạnh mẽ cả về hiệu năng mã hóa lẫn giảm thiểu độ phức tạp thuật toán và mức tiêu thụ năng lượng. Việc kết hợp giữa chuẩn mã hóa truyền thống

H.265/HEVC với phương pháp mã hóa video phân tán DVC là hoàn toàn phù hợp với các ứng dụng truyền thông đa phương tiện. Đặc biệt là các ứng dụng đòi hỏi hiệu năng mã hóa cao, thời gian mã hóa ngắn và sự nhạy cảm thấp đối với các nhiễu kênh truyền như trong mạng video cảm biến điều khiển giao thông, y tế hay mạng lưới các thiết bị di động thông minh.

2.3. Phân tích đặc tính của video giám sát

Ngày nay, hệ thống giám sát video được sử dụng rộng rãi trong rất nhiều ứng dụng quan trọng như bảo vệ cá nhân và an ninh công cộng [11]. Một VSS có thể giám sát và phân tích trong môi trường quan sát theo thời gian thực. Các ứng dụng giám sát video trong thời gian thực thường yêu cầu lưu trữ video mà không bỏ qua bất kỳ phần nào của nội dung trong nhiều tuần hoặc nhiều tháng. Hình 2.2 cho thấy tổng quan của một hệ thống camera giám sát, hệ thống bao gồm 3 phần chính: các nút camera, máy chủ và phía người dùng. Đầu tiên video được quay và xử lý tại nút camera, sau đó được gửi đến máy chủ, luồng bit video tại phía máy chủ có thể được truyền tới người dùng với chất lượng và các độ phân giải khác nhau, tùy vào thiết bị phía người dùng đang sử dụng.



Hình 2.2. Hệ thống camera giám sát

Trong hệ thống giám sát, camera thường được đặt cố định ở một vị trí hoặc di chuyển với góc quay rất nhỏ. Xem xét thực tế này, một số thí nghiệm đã được thực hiện trên các mẫu video huấn luyện khác nhau. Ba video giám sát từ bộ dữ liệu PKU-SVD-A [12, 13] là *Mainroad, Classover* và *Intersection* cùng với một video thường, *BasketballDrill* thu được từ [14] được sử dụng cho thí nghiệm.

Trước tiên, để đánh giá mối tương quan thời gian và chuyển động giữa các khung liên tiếp của video giám sát, sự khác biệt giữa các khung (Frame Difference – FD) được tính:

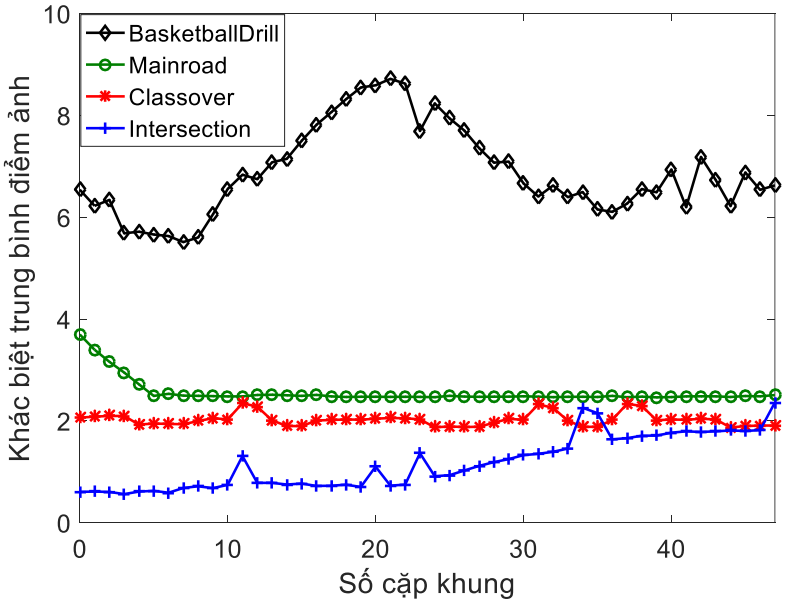
$$FD_t = \sum_{i=1}^{N} |F_t(i) - F_{t+1}(i)| \quad (1)$$

Trong đó t^{th} và i^{th} tương ứng là chỉ số của khung và vị trí điểm ảnh trong mỗi khung F_t, và N là số điểm ảnh trong mỗi khung.

Vì các video huấn luyện có thể có độ phân giải khác nhau, do đó sự khác biệt trung bình điểm ảnh (pixel-averaged difference –PAD) được sử dụng, nó được tính như sau để đánh giá các đặc tính chuyển động:

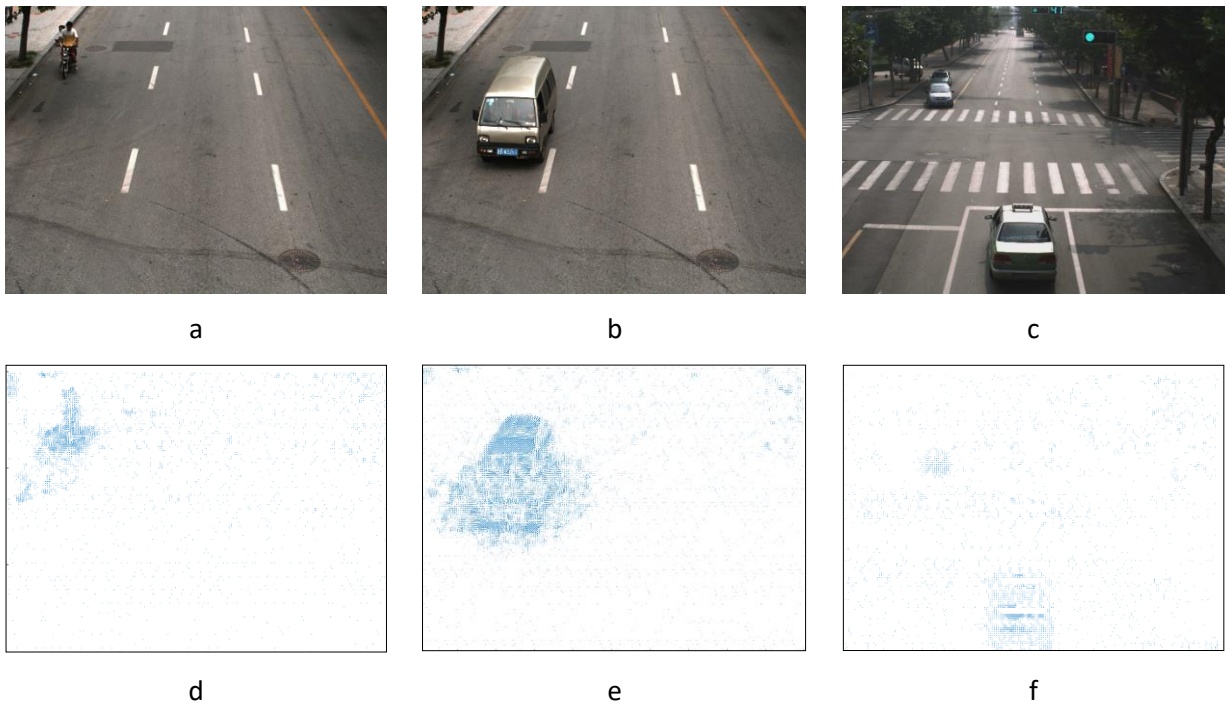
$$PAD_t = \frac{FD_t}{N} \quad (2)$$

Hình 2.3 minh họa *PAD* theo cặp khung liên tiếp của các video thông thường và các video giám sát đã đề cập. Có thể thấy trong hình 2.3, *PAD* giữa các khung trong video giám sát, đáng chú ý là *Mainroad*, *Classover*, và *Intersection* nhỏ hơn rất nhiều so với video tiêu chuẩn, *BasketballDrill*. Trong bối cảnh này, *PAD* nhỏ ngụ ý mối tương quan thời gian cao giữa các khung liên tiếp. Do đó, cần lưu ý rằng các video giám sát thường chứa cảnh tĩnh và ít chuyển động.



Hình 2.3. PAD giữa các khung liên tiếp

Trong thử nghiệm thứ hai, vùng nền trong mỗi khung hình của video giám sát được kiểm tra bằng cách đánh giá trường vectơ chuyển động liên quan đến từng khung hình video. Hình 2.4 minh họa 3 khung được lấy từ video giám sát (a, b, c) và trường vectơ chuyển động tương ứng của chúng (d, e, f).



Hình 2.4. Ví dụ về các khung hình và trường vectơ chuyển động tương ứng trong video giám sát

Như được hiển thị trong hình 2.4, kích thước của vùng chuyển động trong các video giám sát nhỏ hơn nhiều so với diện tích vùng nền. Có thể kết luận rằng trong một video giám sát, các cảnh tĩnh thường chiếm tỉ lệ cao. Do đó, đặc tính quan trọng này được khai thác để xây dựng kiến trúc nén video hiệu quả cho hệ thống giám sát hình ảnh trong luận văn này.

CHƯƠNG 3. GIẢI PHÁP TRÊN SHVC

3.1. Giới thiệu

Chuẩn mã hóa video hiệu suất cao H.265/HEVC là chuẩn mã hóa video mới nhất hiện nay với tên gọi đầy đủ là High Efficiency Video Coding (HEVC) được ủy ban Viễn thông quốc tế ITU-T thông qua và bắt đầu được đưa vào các sản phẩm thương mại hóa. Hiện nay, độ phân giải video đã lên đến 4K/UHD hoặc thậm chí là 8K/UHD. Do đó, H.265/HEVC được sinh ra để nén hiệu quả các video có độ phân giải cực cao này vì ở độ phân giải càng cao thì nó càng thể hiện được lợi thế nén của mình khi so với H.264/AVC. Ưu thế này làm cho độ phân giải video 8K có thể hiển thị và phát trực tuyến thông qua môi trường Internet. SHVC là phần mở rộng của HEVC, hỗ trợ khả năng mở rộng theo thời gian, không gian và chất lượng video.

Trong những năm gần đây, các kích thước mới của tín hiệu video đã bắt đầu xuất hiện. Việc triển khai video độ phân giải cực cao (Ultra High Definition – UHD) có độ dài bit và khoảng màu rộng hơn những video độ nét cao (High Definition – HD) đã thu hút được nhiều sự quan tâm của thị trường. Để cung cấp khả năng tương thích ngược thì khả năng mở rộng về độ dài bit và khả năng mở rộng gam màu cũng được hỗ trợ bởi chuẩn mã hóa SHVC.

Tỷ số tín hiệu trên nhiễu SNR (Signal to Noise Ratio) được sử dụng như một thước đo về chất lượng của tín hiệu và do đó tính chất mở rộng chất lượng cũng được gọi là khả năng mở rộng SNR. Trong mã hóa video, tỷ số tín hiệu cực đại trên nhiễu PSNR (Peak Signal to Noise Ratio) thường được sử dụng làm thước đo SNR.

Trong mã hóa ảnh và video, để đo chất lượng tín hiệu khôi phục của các thuật toán nén có mất mát người ta dùng PSNR. PSNR là một thuật ngữ dùng để tính tỷ lệ giữa giá trị năng lượng tối đa của một tín hiệu với năng lượng nhiễu có ảnh hưởng đến độ chính xác của thông tin. Đơn vị của nó thường được biểu diễn bởi logarit decibel (dB).

$$PSNR = 20.log_{10}(\frac{MAX_I}{\sqrt{MSE}}) \quad (3)$$

Trong đó: MAX_I là giá trị tối đa của điểm ảnh trên ảnh. Khi một điểm ảnh được biểu diễn bởi 8 bit thì giá trị của nó sẽ là 255. Tổng quát, khi tín hiệu được biểu diễn bởi b bit trên một đơn vị lấy mẫu thì MAX_I bằng $2^b – 1$.

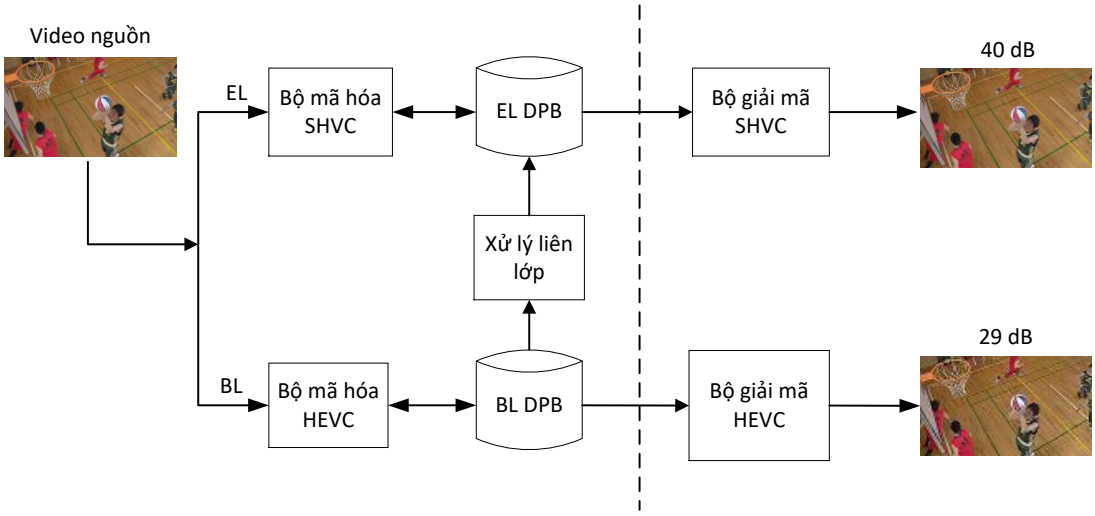
MSE (Mean Squared Error) là sai số toàn phương trung bình được dùng cho ảnh hai chiều có kích thước là MxN. Trong đó Org là ảnh gốc còn Rec là ảnh được khôi phục tương ứng. MSE được tính qua công thức:

$$MSE = \frac{1}{MxN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[Org(i,j) - Rec(i,j)]^2 \quad (4)$$

Giá trị thông thường của PSNR trong nén ảnh và video có mất mát sẽ nằm từ 30 đến 50 dB, giá trị PSNR càng cao thì càng tốt vì nó sẽ mang lại chất lượng hình ảnh tốt hơn. Giá trị có thể chấp nhận được khi truyền tín hiệu không dây có tổn hao vào khoảng từ 20 – 25 dB.

3.2. Cấu trúc SHVC

Như hình 3.1 minh họa, SHVC bao gồm: Lớp cơ bản (BL – Base Layer) chứa các thông tin cơ bản nhất để đảm bảo chất lượng tối thiểu để bộ giải mã có thể giải mã được. Lớp nâng cao (EL – Enhancement Layer) được tạo ra nhằm nâng cao chất lượng về tỷ lệ khung hình, độ phân giải và tái tạo độ nét khi được nội suy từ lớp cơ bản BL. Bởi vì tín hiệu video đầu vào cho các lớp cơ bản và nâng cao có cùng nội dung nên sự tương quan giữa các lớp này là rất lớn. Do đó, số lượng bit (bit rate) cần thiết để mã hóa lớp nâng cao có thể được giảm đáng kể nếu ta có thể khai thác được sự tương quan này. Ví dụ, trong các chuẩn nén H.264/SVC, số bit cần để mã hóa lớp nâng cao được giảm bằng thuật toán dự đoán liên lớp ILP (Inter Layer Prediction) với ba chế độ cơ bản là dự đoán kết cấu liên lớp (Texture), sự đoán chuyển động liên lớp (Motion) và dự đoán tín hiệu dư thừa (Residual). Trái ngược với SVC, SHVC đã áp dụng cấu trúc mã hóa vòng kín ở mỗi lớp nén và do đó, chỉ có thể thay đổi phần tử cú pháp mức cao (High Level Syntax - HLS) để nâng cấp từ giải pháp HEVC lên SHVC. Theo cách tiếp cận HLS, mô đun xử lý liên lớp được thêm vào để liên kết lớp cơ bản với các lớp nâng cao. Trong mô-đun này, thông tin về kết cấu và chuyển động có nguồn gốc từ các lớp BL hoặc thấp hơn sẽ được tiến hành để sử dụng tối ưu tại các lớp EL.



Hình 3.1. Cấu trúc SHVC 2 lớp

Có thể thấy trong cấu trúc mã hóa video SHVC ở Hình 3.1, video nguồn được chia thành 2 lớp riêng biệt, trong khi lớp BL được gửi đến bộ mã hóa HEVC thì EL được gửi đến bộ mã hóa SHVC. Để đạt được hiệu quả dự đoán lớp EL, việc xử lý liên lớp được áp dụng cho các hình ảnh BL đã giải mã trong bộ đệm hình ảnh đã giải mã (Decoded Picture Buffer - DPB) BL, sau khi hình ảnh được xử lý sẽ được đưa vào EL DPB. Chất lượng sau giải mã của lớp BL dùng bộ mã hóa và giải mã HEVC, chỉ đạt 29 dB trong khi dùng SHVC để mã hóa và giải mã cho lớp EL, có sử dụng thông tin của lớp BL và mô-đun xử lý liên lớp, chất lượng đã tăng lên 40 dB.

Ngoài việc mã hóa sử dụng một bộ mã hóa đơn lớp HEVC thì SHVC còn có thể sử dụng các bộ mã hóa như H.264/AVC, MPEG4 hoặc thậm chí là các bộ codec khác để mã hóa và giải mã, do đó nó còn có thêm khả năng mở rộng giữa các bộ codec với nhau. Điều này cho thấy được sự linh hoạt của chuẩn mã hóa khả chuyển SHVC so với các chuẩn mã hóa không khả chuyển khác. Sau khi giải mã, hình ảnh EL tái tạo được cung cấp cho bộ giải mã SHVC cùng với một số thông tin liên quan đến hình ảnh BL.

Hình 3.1 chỉ mô tả hệ thống mã hóa video SHVC với hai lớp. Tuy nhiên, ta có thể sử dụng chuẩn mã hóa SHVC để mở rộng ra thêm nhiều lớp hơn. Khi thêm nhiều lớp, chỉ có BL có thể được cung cấp từ các chuẩn mã hóa khác và tất cả các lớp còn lại phải được gửi trong băng tần như một phần dòng bit SHVC. Bên cạnh đó, việc xử lý lớp liên kết có thể được áp dụng trên các hình ảnh tái tạo từ bất kỳ lớp tham chiếu nào mà lớp hiện tại tham chiếu tới. Bằng cách giữ lại cách thức hoạt động của bộ mã hóa video một lớp như bộ HEVC do đó thời gian mã hóa của SHVC có thể được giảm đi đáng kể. Để đạt được dự đoán liên lớp, thông tin BL duy nhất mà EL cần truy cập đó là hình ảnh được tái tạo từ DPB của BL.

3.3. Phương pháp mã hóa liên quan

Trong mã hóa video, một nhóm các hình ảnh GOP (Group of Picture) chỉ định thứ tự các khung nội ảnh (intra) và các khung liên ảnh (inter) được sắp xếp. GOP là tập hợp các hình ảnh liên tiếp trong một luồng video được mã hóa. Bởi vì các khung hình liên tiếp trong một video đều có sự tương quan về thời gian và không gian nên do đó các khung hình trong một GOP sẽ có sự tương quan với nhau rất lớn. Mỗi luồng video được mã hóa bao gồm các GOP liên tiếp nhau từ đó khung hiển thị được tạo ra. Một GOP mới được tạo ra trong quá trình nén thì không cần bất kỳ những khung hình ngoài GOP đó để mã hóa cho nó do đó mỗi GOP có tính độc lập về mã hóa.

Một GOP có thể chứa các loại khung hình sau:

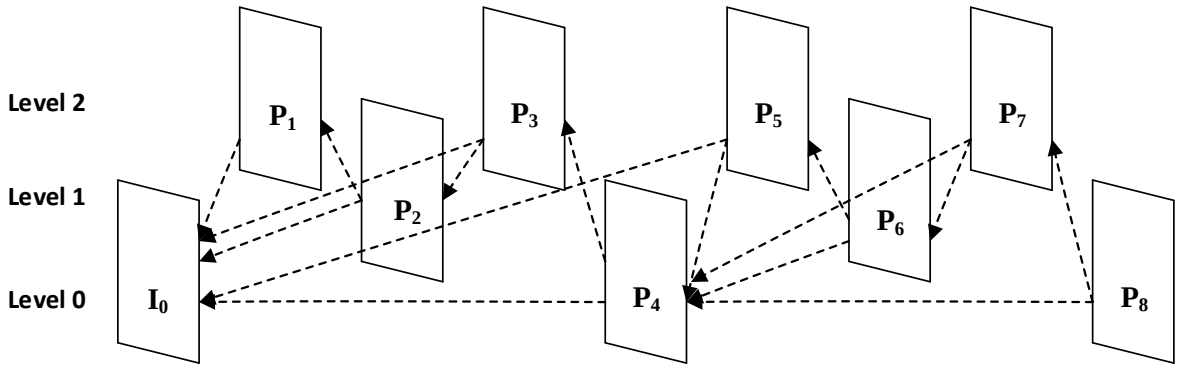
- Khung hình I (Mã hóa trong khung – Intra coded pictures) là một hình ảnh được mã hóa độc lập, không phụ thuộc vào các hình ảnh khác. Đây là khung hình đầu tiên trong mỗi GOP và luôn có chất lượng cao nhất vì được mã hóa ở mức lượng tử hóa QP (Quantization Parameter) là thấp nhất. Nó có chất lượng cao nhất là do nó sẽ được dùng để làm tham chiếu đến các khung hình khác trong GOP để mã hóa các khung hình còn lại đó.

- Khung hình P (Mã hóa dự đoán hình ảnh – Predictive coded pictures) chứa thông tin sai khác của chuyển động liên quan đến hình ảnh được giải mã trước đó.

- Khung hình B (Bi-predictive coded pictures) làm việc giống với khung hình loại P và sử dụng các khung hình I và P làm tham chiếu.

Theo thứ tự trong một GOP thì khung hình I có chất lượng cao nhất vì nó là khung hình gốc được dùng để mã hóa các khung hình khác trong GOP và sau đó đến các khung hình P và B.

Chương này trong luận văn sẽ tập trung khai thác kiểu mã hóa theo Low-delay P (LDP). Trong cấu trúc mã hóa này thì hình ảnh đầu tiên được mã hóa dưới dạng khung I và các hình ảnh tiếp theo sẽ được mã hóa dưới dạng khung P. Do chế độ low-delay không cho phép sắp xếp lại thứ tự mã hóa do đó độ trễ trong chế độ mã hóa này là rất nhỏ vì các khung hình được mã hóa liên tiếp nhau do đó chế độ mã hóa này phù hợp với mã hóa các video online hơn là offline. Trong low-delay, ở level càng thấp thì càng cần có chất lượng cao và chúng được mã hóa với chỉ số QP thấp hơn các level cao hơn.



Hình 3.2. Cấu trúc dự đoán theo low-delay P

Hình 3.2 mô tả cấu trúc dự đoán trong low-delay P. Chiều mũi tên sẽ là chiều tham chiếu tới. Nó sẽ tận dụng những đặc điểm giống nhau về mặt thời gian giữa các khung hình để tăng hiệu quả nén.

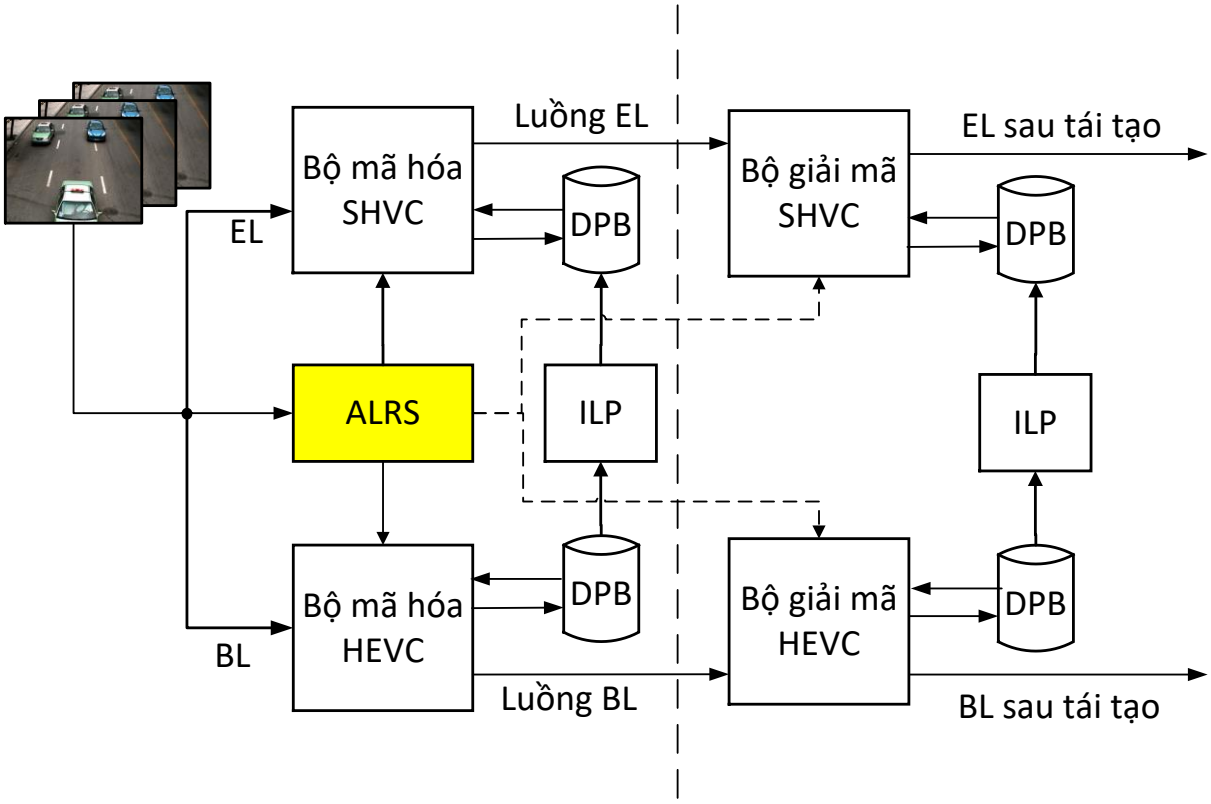
Bởi vì hầu hết các video giám sát là các video thời gian thực, do đó trong khuôn khổ luận văn này sẽ chủ yếu tập trung vào sử dụng kiểu mã hóa low-delay vì nó có độ trễ thấp.

3.4. Đề xuất trên SHVC

Mặc dù SHVC là tiêu chuẩn mã hóa video khả chuyển mới nhất nhưng nó vẫn không phải là một giải pháp nén tối ưu cho các hệ thống giám sát vì nó chủ yếu được thiết kế cho nội dung video thông thường. Trong một hệ thống giám sát hình ảnh, camera thường được đặt ở một vị trí cố định hoặc di chuyển với góc rất hẹp; do đó, nội dung video giám sát thường chứa diện tích nền lớn cũng như có mối tương quan thời gian cao giữa các khung. Để khai thác các đặc điểm này, luận văn này đề xuất một giải pháp nén SHVC cải tiến được thiết kế cho nội dung video giám sát.

3.4.1. Kiến trúc SSVC

Giải pháp mã hóa video khả chuyển cho giám sát (Surveillance Scalable Video Coding - SSVC) được tạo ra dựa trên: i) cấu trúc SHVC thông thường; ii) cơ chế cập nhật và lựa chọn tham chiếu dài hạn thích ứng.



Hình 3.3. Kiến trúc SSVC

Hình 3.3 minh họa kiến trúc SSVC được đề xuất trong luận văn, trong đó chỉ thể hiện nén và giải nén cho một lớp BL và một lớp EL. Cơ chế cập nhật và lựa chọn tham chiếu dài hạn thích ứng (Adaptive Long-term Reference Selection - ALRS) được làm nổi bật trong kiến trúc SSVC ở hình 3.3. Phương pháp này đề xuất sử dụng khung tham chiếu đầu tiên trong GOP làm khung tham chiếu dài hạn cho các khung tiếp theo của chuỗi video. Việc sử dụng khung đầu tiên trong GOP làm khung tham chiếu để mã hóa cho các khung tiếp theo là do khung đầu tiên trong mỗi GOP luôn có chất lượng cao nhất vì được mã hóa bởi chỉ số lượng tử hóa QP là thấp nhất. Khung này có thể được sử dụng làm tham chiếu cho bất kỳ các GOP tiếp theo mà chuẩn mã hóa SHVC không có. Kỹ thuật này được áp dụng giống nhau đối với cả hai lớp BL và EL. Đây là một kỹ thuật đơn giản nhưng nó mang lại hiệu quả mã hóa tương đối tốt.

Video giám sát sau khi được ghi lại từ cảm biến camera, được xử lý bằng giải pháp SSVC theo các bước chính sau:

Bước 1. *Adaptive Long – Term Reference Selection (ALRS):* Mô-đun này định nghĩa khung tham chiếu phù hợp cho dự đoán liên khung SSVC. Nếu khung được chọn sẽ được đánh chỉ số là khung tham chiếu dài hạn và được lưu trong bộ đệm ảnh đã giải

mã, cần một cờ mã hóa để báo hiệu thông tin này để đảm bảo phía giải mã cũng biết khung được chọn.

Bước 2. *Mã hóa BL:* Sau khi định nghĩa chỉ số khung tham chiếu dài hạn, khung BL được mã hóa sử dụng chuẩn HEVC thông thường. Cấu trúc và chuyển động của nó được lưu ở bộ đệm ảnh giải mã (DPB) để khai thác sau này cho mã hóa các khung EL.

Bước 3. *Mã hóa EL*: Ở bước này, thông tin BL đã giải mã sẽ được xử lý bởi mô-đun xử lý liên lớp [6]. Cùng với các khung tham khảo thời gian dài, tham khảo thông tin từ BL để nén thông tin EL tốt hơn.

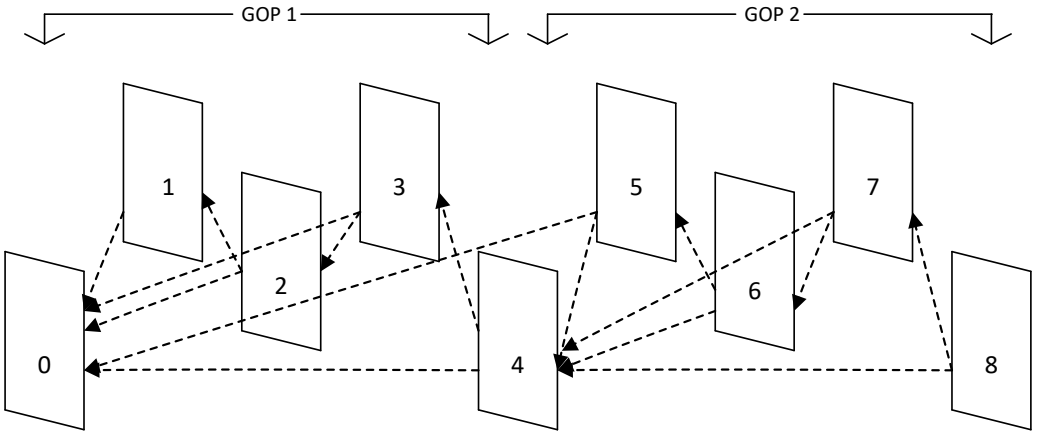
Bước 4. Cả luồng BL và EL được gộp lại để gửi tới phía giải mã.

3.4.2. Cấu trúc tham chiếu thời gian dài

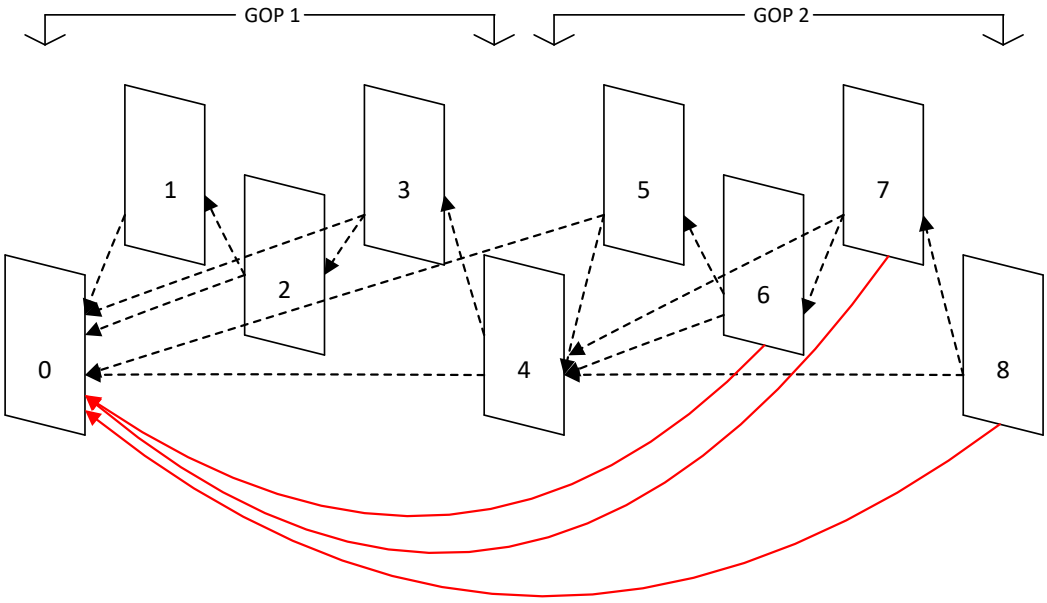
Để làm rõ cấu trúc tham chiếu dài hạn được đề xuất sử dụng trong SSVC, cấu trúc mã hóa độ trễ thấp (low-delay - LD), thường được sử dụng trong các tiêu chuẩn HEVC và SHVC [4] được so sánh như trong Hình 3.4. Cần lưu ý rằng cấu trúc tham chiếu dài hạn được sử dụng cho cả hai lớp cơ bản và nâng cao.

Như được hiển thị trong hình 3.4, ở chuẩn mã hóa video thông thường, một khung hình có thể được dùng bởi tối đa 5 khung hình liên tiếp khác, tức là số khung 1, 2, 3, 4 và 5 có thể tham chiếu thông tin được giải mã của khung số 0. Tuy nhiên, đối với đề xuất tham chiếu dài hạn thích nghi, khung số 6, 7, 8 hoặc hơn thế cũng có thể tham chiếu đến khung số 0. Điều này có thể cho phép khai thác mối tương quan thời gian cao giữa các khung trong video giám sát.

Theo các phân tích ở trên, sự thay đổi nền của các video giám sát là tương đối thấp. Vì vậy việc tận dụng lợi thế của khung đầu tiên trong GOP như là một khung tham chiếu dài hạn sẽ giúp tận dụng được các đặc tính riêng biệt của các video giám sát qua đó nâng cao hiệu suất mã hóa cho toàn bộ video. Bởi vì nền cũng sẽ thay đổi theo thời gian nên khung tham chiếu dài hạn LTR cũng cần phải được cập nhật để cải thiện hiệu suất mã hóa. Cơ chế cập nhật này sao cho tối ưu sẽ được trình bày trong phần tiếp theo của luận văn.



(a)



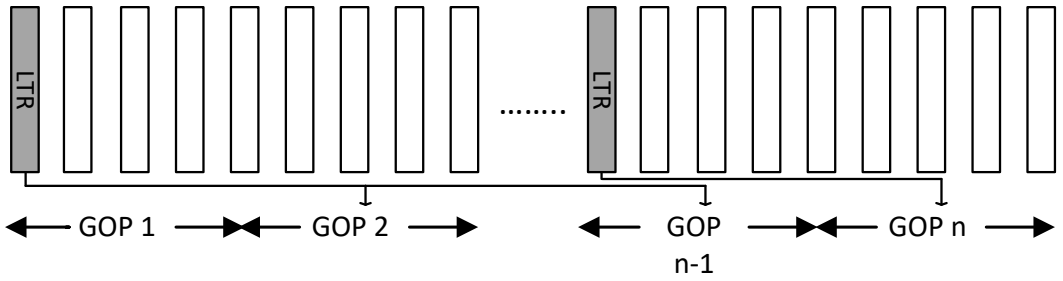
(b)

Hình 3.4. Cấu trúc dự đoán độ trễ thấp (LD) trong chuẩn mã hóa video thông thường (a) và trong SSVC (b)

3.4.3. Lựa chọn và cập nhật khung tham chiếu thời gian dài thích nghi

Trong video giám sát, khi một đối tượng chuyển động thì một số cảnh bị thay đổi. Trong trường hợp này, chỉ dùng duy nhất một khung tham khảo dài hạn thì sẽ không hiệu quả. Xem xét vấn đề này, luận văn đề xuất giải pháp cập nhật thích nghi khung tham khảo dài hạn.

Hình 3.5 minh họa một ví dụ về cơ chế cập nhật khung tham chiếu dài hạn trong đó khung tham chiếu dài hạn mới được cập nhật dựa trên đặc tính nội dung của video giám sát.



Hình 3.5. Giải pháp ALRS

Xét từ GOP thứ nhất đến GOP thứ n-1, nếu sự thay đổi không quá khác biệt, xét qua thông số tổng trị tuyệt đối sai khác (Sum of Absolute Differences - SAD) là một thước đo tính tương quan giữa các hình ảnh, nếu SAD nhỏ hơn ngưỡng nhất định thì khung tham chiếu đầu tiên của GOP 1 vẫn được sử dụng cho GOP n-1. Tiếp theo, đến GOP thứ n, nếu khung hình đầu tiên của GOP n khi so với khung tham chiếu có SAD lớn hơn ngưỡng thì LTR sẽ được thay thế bằng khung hình đầu tiên của GOP n-1. Lý do là các khung hình gần nhau thì sẽ có sự tương quan lớn về mặt thời gian do đó khung hình đầu tiên của GOP được chọn. Điều này sẽ được lặp lại tương tự cho các GOP tiếp theo cho mã hóa chuỗi video. Việc lựa chọn ngưỡng sao cho phù hợp sẽ được nói dưới đây. Một thuật toán lựa chọn tham chiếu dài hạn đơn giản được đề xuất sử dụng tổng chênh lệch tuyệt đối (SAD) giữa khung được mã hóa hiện tại, F_t và khung tham chiếu dài hạn của nó, LTR_t.

SAD được đo bằng công thức:

$$SAD_t = \sum_{i=0}^{N-1} |F_t(i) - LTR_t(i)| \quad (5)$$

Ở đây, i^{th} là chỉ số của điểm ảnh của khung có N điểm ảnh trong khi t^{th} là chỉ số của khung hình trong chuỗi video giám sát. F_t là khung đầu tiên của mỗi GOP trong video giám sát và LTR_t là khung tham chiếu dài hạn gần nhất được sử dụng.

Để đánh giá mối tương quan giữa khung đầu mỗi GOP và khung tham chiếu của nó, sử dụng ngưỡng thích nghi, ngưỡng ban đầu được khởi tạo như sau:

$$ThD_0 = \frac{1}{8}\sum_{j=0}^{7}|SAD_j| \quad (6)$$

Vì các khung hình trong một GOP có tính tương quan về mặt thời gian cao và do chỉ sau ít nhất 2 GOP thì LTR mới được cập nhật LTR mới nên giá trị sai khác trung bình của 2 GOP này sẽ được chọn là ngưỡng khởi tạo để cập nhật LTR cho các GOP tiếp theo. Trong cơ chế lựa chọn khung tham chiếu dài hạn, nếu $(SAD_t > ThD_t)$, khung tham chiếu gần nhất sẽ được cập nhật là khung tham chiếu dài hạn cho khung hiện tại và các khung liên tiếp, ngưỡng mới được cập nhật bằng giá trị SAD mới, ngược lại nếu $(SAD_t \leq ThD_t)$, vẫn sử dụng khung tham chiếu dài hạn hiện tại.

3.5. Kết quả giải pháp SSVC

Để đánh giá giải pháp SSVC được đề xuất trong chương này, bốn video giám sát phổ biến thu được từ bộ dữ liệu PKU-SVD-A [13] được sử dụng cho thí nghiệm. Bảng 1 trình bày tên và đặc điểm trong khi hình 3.6 minh họa các khung đầu tiên của các video được sử dụng.

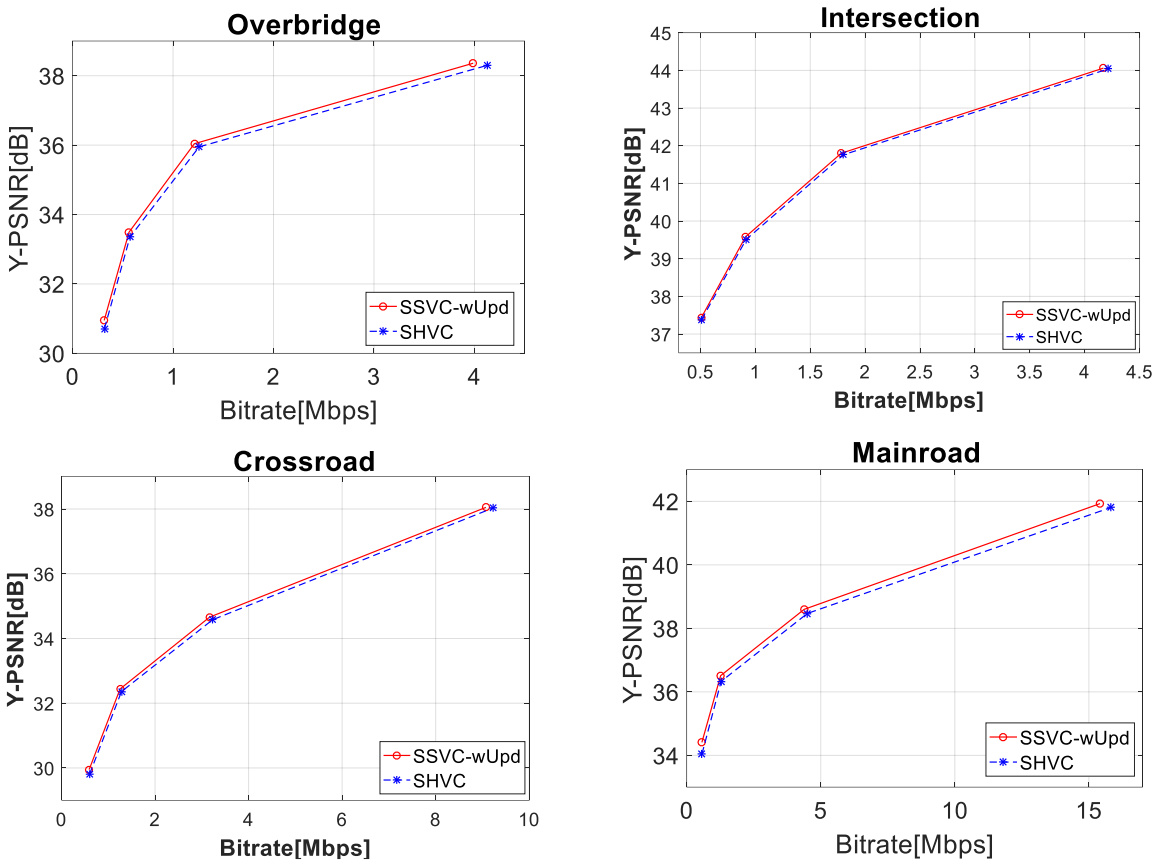
Crossroad Intersection Mainroad Overbridge

Hình 3.6. Khung đầu tiên của các video thử nghiệm

Bảng 1. Tổng hợp các điều kiện kiểm tra SSVC

Tên và độ phân giải các video tương ứng	1. *Crossroad*, 720×576 2. *Intersection*, 1600×1200 3. *Mainroad*, 1600×1200 4. *Overbridge*, 720×576
Tốc độ khung hình và số khung hình được nén	@30Hz, 297 khung
Nhóm ảnh (Group of Pictures - GOP)	Low_delay_P (IPPP…) (GOP4)
Tham số lượng tử (Quantization Parameters -QP)	QP_B = {38, 34, 30, 26) $QP_E = QP_B - 4$
Phần cứng	- Processor: Intel® Core i7 - 3.2 GHz - RAM: 8.00 GB - System: Win 10, 64-bit - Environment: Microsoft Visual Studio 2017

Hình 3.7 minh họa RD performance trong khi Bảng 2 thể hiện kết quả tiết kiệm BD-Rate khi so sánh giải pháp SSVC đề xuất với chuẩn SHVC cho 4 chuỗi video giám sát.



Hình 3.7. So sánh RD performance với trường hợp SSVC có cập nhật khung tham chiếu dài hạn

Bảng 2. Tiết kiệm BD-Rate với SSVC

Chuỗi video	**SSVC-woUpd vs. SHVC**	**SSVC-wUpd vs. SHVC**
Crossroad	-1,08	-4,56
Overbridge	-2,57	-6,15
Mainroad	-1,79	**-8,56**
Intersection	-0,69	-2,24
Trung bình	**-1,53**	**-5,38**

Như thể hiện trong Bảng 2, cả SSVC-woUp (không cập nhật LTR) và SSVC-wUp (có cập nhật LTR) đều thu được hiệu năng nén tốt hơn khi so với chuẩn SHVC, trung bình BD-rate saving là 1,53% và 5,38% tương ứng. Giải pháp dùng LTR trong chương này thu lại kết quả tốt cho tất cả video được kiểm thử, đặc biệt là video *Mainroad* với ít chuyển động và đối tượng đơn.

CHƯƠNG 4. GIẢI PHÁP TRÊN DSVC

Một số chuẩn mã hóa video như HEVC [2] và phần mở rộng của nó, SHVC [6] chủ yếu được thiết kế cho các video thông thường. Xem xét đặc tính tĩnh của các video giám sát, các tác giả trong [15, 16] đã đưa ra mô hình nền dựa trên dự đoán thích nghi cho nén video giám sát. Sau đó có một lượng lớn các cải tiến mã hóa video cho hệ thống giám sát được đưa ra [17 - 19]. Tuy nhiên, vì cấu trúc mã hóa video giám sát thường được phát triển dựa trên các chuẩn mã hóa video dự đoán thông thường, ví dụ: H.264/AVC [4] hoặc HEVC [2], hiệu năng mã hóa của nó thường đi cùng với mức độ tính toán cao, vì thế phía mã hóa rất phức tạp. Trong trường hợp này, yêu cầu độ phức tạp mã hóa thấp cho hệ thống video giám sát có thể không được thỏa mãn. Ngoài ra, các giải pháp mã hóa video giám sát trước đó [15 - 19] không khả chuyển vì chỉ mã hóa một lớp.

Trong bối cảnh này, xem xét nhu cầu của một giải pháp mã hóa video mạnh mẽ thường đòi hỏi hiệu quả nén cao, khả năng mở rộng và độ phức tạp thấp, luận văn đã đề xuất một giải pháp mã hóa video có thể mở rộng mới, được thiết kế đặc biệt cho dữ liệu video giám sát. Đề án mã hóa video có thể mở rộng cho giám sát được phát triển dựa trên sự kết hợp của các tiêu chuẩn mã hóa video dự đoán truyền thống, HEVC và SHVC với mô hình mã hóa video phân tán mới nổi [20]. Khi cấu trúc mã hóa lớp được thông qua, giải pháp mã hóa video phân tán khả chuyển cho hệ thống giám sát được đề xuất, cụ thể là (Surveillance – Distributed Video Coding S-DSVC), có thể cung cấp các tính năng mở rộng theo thời gian và chất lượng. Ngoài ra, một số công cụ mã hóa cũng được giới thiệu để tăng thêm hiệu suất nén của giải pháp S-DSVC được đề xuất.

4.1. Cấu trúc DVC

Để hiểu hơn về DSVC, trước hết mã hóa nguồn phân tán được giới thiệu vì đây là nền tảng cho mã hóa video phân tán. Mã hóa nguồn phân tán dựa trên hai kết quả lý thuyết thông tin quan trọng là định lý Slepian-Wolf và Wyner-Ziv.

Distributed video coding (DVC) là một cách tiếp cận mã hóa khác, nhắm vào độ phức tạp mã hóa thấp tại phía mã hóa và tránh lỗi tại phía giải mã [20]. DVC được phát triển từ 2 định lý thông tin nổi tiếng, Slepian-Wolf [21] and Wyner-Ziv [22]. Đã có sự chú ý lớn về DVC trong những thập kỷ gần đây với nhiều đóng góp đáng kể, đáng chú ý là cả kiến trúc mã hóa thực tế và cải tiến các công cụ mã hóa [23, 24]. Trong DVC, mối tương quan thời

gian chủ yếu được khai thác ở phía bộ giải mã bằng cách tạo thông tin phụ trợ (Side Information – SI) [25] trong khi phía bộ mã hóa được thiết kế theo cách rất đơn giản. Do đó, giải pháp mã hóa này rất hấp dẫn đối với các ứng dụng mã hóa video mới nổi, ví dụ: mạng cảm biến hình ảnh, hệ thống giám sát và viễn thám. Các nghiên cứu gần đây cũng chỉ ra rằng DVC thường phù hợp để mã hóa các video có nội dung chuyển động thấp và tĩnh [20, 26]. Như được đánh giá trong [27], giải pháp mã hóa thực tế DVC yêu cầu độ phức tạp mã hóa thấp hơn nhiều so với các tiêu chuẩn mã hóa video dự đoán truyền thống, ví dụ, H.264/AVC hoặc HEVC trong khi cung cấp khả năng phục hồi lỗi mạnh mẽ hơn và mã hóa video hiệu quả hơn.

Định lý Slepian và Wolf (1973) nói rằng tốc độ tối thiểu để mã hóa hai nguồn tương quan, X và Y giống như tốc độ tối thiểu cho joint encoding, điều này có nghĩa là joint entropy $H(X,Y)$, với xác suất xuất hiện lỗi nhỏ tùy ý cho các chuỗi dài, với điều kiện là mối tương quan của chúng được biết đến tại phía mã hóa và giải mã. Định lý này rất quan trọng vì đây là lần đầu tiên thiết lập ranh giới tỷ lệ cho một mã hóa riêng biệt nhưng giải mã chung của hai nguồn tương quan như được trình bày trong các bất đẳng thức sau:

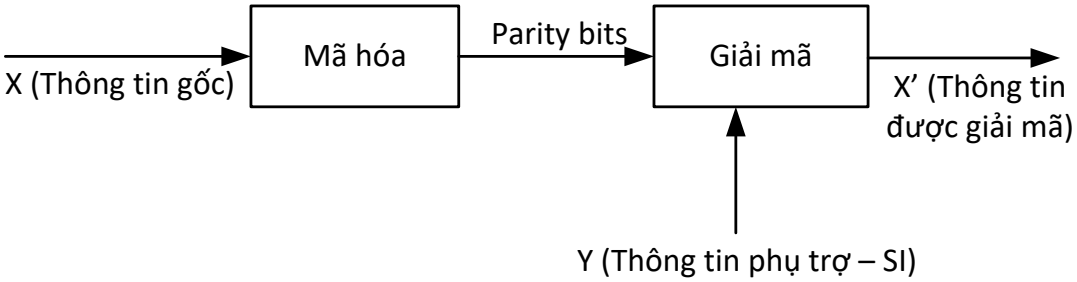
$$R_X \geq H(X \mid Y), R_Y \geq H(Y \mid X) \qquad (7)$$
$$R_X + R_Y \geq H(X,Y)$$

ở đây $H(Y \mid X)$ biểu thị entropy có điều kiện và $H(X,Y)$ biểu thị joint entropy của nguồn X và Y.

Tuy nhiên định lý Slepian và Wolf chỉ đề cập đến trường hợp mã hóa không tổn thất mà không xem xét đối với các giải pháp mã hóa video thực tế do tỷ lệ nén thấp liên quan. Năm 1976, Wyner và Ziv [22] đã mở rộng định lý Slepian và Wolf sang trường hợp nén có tổn thất. Định lý Wyner và Ziv nói rằng, đối với nguồn X có thông tin phụ trợ là Y có sẵn ở bộ giải mã, tốc độ cần thiết để đạt được một biến dạng nhất định khi một số thông tin bên có sẵn tại bộ giải mã chỉ tuân theo $R_{(X|Y)}(D) \leq R_{(X|Y)}^{WZ}(D)$ trong đó $R_{(X|Y)}(D)$ là tốc độ thu được khi SI có sẵn ở cả bộ mã hóa và bộ giải mã. Do đó, khi phụ thuộc thống kê chỉ được khai thác tại bộ giải mã, tốc độ tối thiểu để truyền X ở cùng một biến dạng D có thể tăng hoặc bằng so với trường hợp phụ thuộc thống kê được khai thác ở cả bộ mã hóa và bộ giải mã (thường được áp dụng trong các tiêu chuẩn mã hóa video, ví dụ, H.264/AVC và HEVC).

Hai định lý này đã mở ra các cơ hội mới cho cho mã hóa video hay còn gọi là mã hóa video phân tán. Do việc mã hóa được thực hiện độc lập và việc giải mã thực hiện kết hợp nên dư thừa thời gian chỉ được khai thác tại bộ giải mã thay vì bộ mã hóa, do đó bộ mã hóa có độ phức tạp rất thấp nếu so với mã hóa video truyền thống. Trong DVC, chuỗi X là thông

tin gốc, thông tin được mã hóa Wyner-Ziv và chuỗi Y là thông tin phụ trợ (SI). Thông tin phụ trợ là một dự đoán của thông tin chính và được tạo ra ở phía giải mã. Bộ giải mã sẽ sử dụng thông tin phụ trợ này kết hợp với một số thông tin bổ sung từ phía mã hóa gửi đến để dự đoán ra thông tin chính X (Hình 4.1). Do đó, nếu tương quan giữa chuỗi X và chuỗi Y càng cao thì số lượng bit bộ mã hóa cần gửi sang cho bộ giải mã càng ít. Nói cách khác, hiệu quả các kỹ thuật tạo thông tin phụ trợ sẽ ảnh hưởng đáng kể đến hiệu năng tổng thể của codec video phân tán hay còn gọi là codec video Wyner-Ziv (WZ).

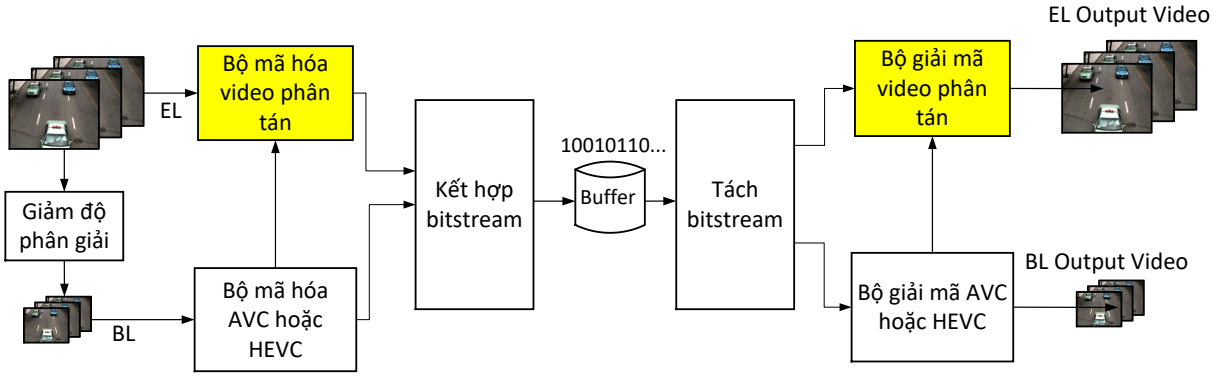


Hình 4.1. Phương pháp nén video phân tán với thông tin phụ trợ

4.2. Cấu trúc DSVC

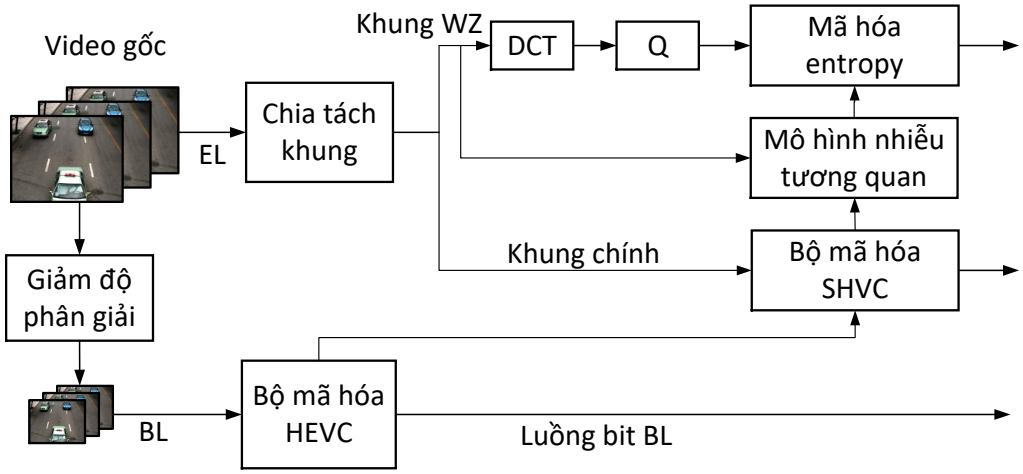
DSVC để chỉ tất cả các giải pháp mã hóa video mà ở đó có sử dụng các nguyên tắc mã hóa video phân tán đồng thời vẫn khả chuyển. Cũng giống như các hệ thống mã hóa video liên lớp khác, DSVC gồm một lớp cơ bản (BL) và một hoặc nhiều lớp nâng cao (EL).

Nhóm các khung hình (GOP) được thiết lập bằng 2, nghĩa là cứ một khung hình WZ nằm giữa hai khung hình chính. Một trong những tính năng chính của mã hóa video có thể mở rộng thế hệ mới là lớp cơ bản phải tương thích với không chỉ H.264/AVC trước đó mà cả các chuẩn H.265/HEVC như minh họa trong Hình 4.2 bên dưới. Ở phương pháp mã hóa video DSVC này, video nguồn được chia thành 2 lớp BL và EL. Tại phía mã hóa, để có thể tương thích với các chuẩn mã hóa video hiện nay, thông thường lớp cơ bản được mã hóa bằng các chuẩn mã hóa video truyền thống như H.264/AVC hay HEVC và lớp nâng cao được mã hóa theo nguyên tắc mã hóa video phân tán hay còn gọi là mã hóa Wyner-Ziv (WZ). Sau đó các luồng bitstream của EL và BL được kết hợp và chuyển đến bộ đệm. Ở phía giải mã, thông tin trong bộ đệm được tách thành luồng BL và EL, sau đó giải mã bằng các chuẩn tương ứng như khi mã hóa.



Hình 4.2. Cấu trúc DSVC

A. Quá trình mã hóa



Hình 4.3. Bộ mã hóa DSVC

Hình 4.3 thể hiện quá trình mã hóa trong DSVC. Đầu tiên, chuỗi video gốc sẽ được đưa qua module giảm độ phân giải để tạo ra các khung hình có độ phân giải không gian nhỏ hơn cho lớp cơ bản. Các khung hình lớp cơ bản sẽ được mã hóa bằng bộ mã hóa HEVC. Đối với lớp nâng cao, quá trình mã hóa như sau:

- Chia tách khung: Các khung EL được chia thành các khung WZ (WZ frame) và khung chính (Key frame) xen kẽ nhau. Các khung hình lẻ là các khung chính còn các khung chẵn là các khung WZ.

- Bộ mã hóa SHVC: Các khung chính được mã hóa bằng cách sử dụng bộ mã hóa SHVC, chuẩn mở rộng khả chuyển của chuẩn mã hóa video HEVC.

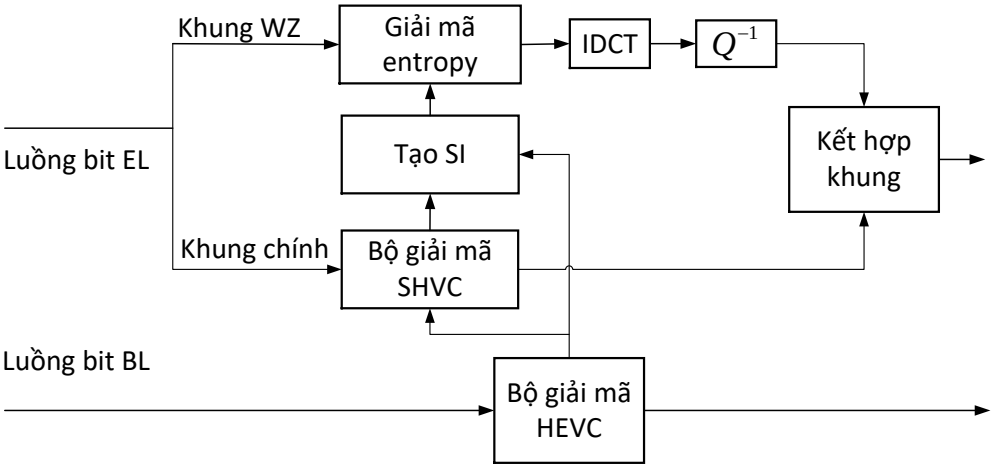
- Mô hình nhiễu tương quan: Phương pháp mã hóa video WZ thực hiện mã hóa sự sai khác giữa thông tin gốc tại bộ mã hóa và thông tin phụ trợ được tạo ra tại phía giải mã. Do đó, cả bộ mã hóa và bộ giải mã cần phải biết về tương quan thống kê giữa thông tin gốc và thông tin phụ trợ. Vì thế khối này thực hiện mô hình hóa nhiễu tương quan giữa khung hình WZ gốc và khung hình thông tin phụ trợ tương ứng. Thông tin đầu ra sẽ được gửi tới bộ mã hóa entropy để mã hóa cho khung hình WZ.

- DCT: Để khai thác dư thừa không gian, các khung hình WZ sẽ được biến đổi DCT để tạo ra các ma trận hệ số tương ứng.

- Lượng tử hóa: Để loại bỏ dư thừa không gian và tận dụng các đặc điểm của thị giác người, các ma trận hệ số DCT được đưa qua bộ lượng tử hóa với các hệ số lượng tử tùy theo chất lượng mong muốn.

- Mã hóa entropy: Bộ mã hóa này thực hiện mã hóa các hệ số DCT đã lượng tử hóa có sử dụng thông tin từ khối mô hình hóa nhiễu tương quan. Bộ mã hóa này có thể là bộ mã hóa kênh hoặc bộ mã hóa syndrome tùy thuộc vào các cách tiếp cận khác nhau.

B. Quá trình giải mã



Hình 4.4. Bộ giải mã DSVC

Kiến trúc bộ giải mã DSVC được mô tả trong Hình 4.4. Trước hết, luồng bit cơ sở được giải mã bởi bộ giải mã HEVC. Luồng bit nâng cao được chia thành các khung hình chính và khung hình WZ. Các khung hình chính được giải mã bởi bộ giải mã SHVC. Các khung hình WZ được giải mã như sau:

- Tạo SI: Thông tin phụ trợ được coi như một phiên bản nhiễu của thông tin gốc. Nó được tạo ra dựa trên các thông tin sẵn có ở phía giải mã. Trong kiến trúc DSVC liên lớp không gian, thông tin phụ trợ có thể được xây dựng dựa trên các khung hình chính trước và sau và các khung hình tương ứng ở lớp cơ bản.

- Giải mã entropy: thực hiện giải mã thông tin gốc bằng cách sửa sai thông tin phụ trợ dựa trên các thông tin nhận được từ phía mã hóa gửi tới.

- Q^{-1} và IDCT: Sau khi giải mã entropy, thông tin được tái tạo bằng cách giải lượng tử và biến đổi DCT ngược.

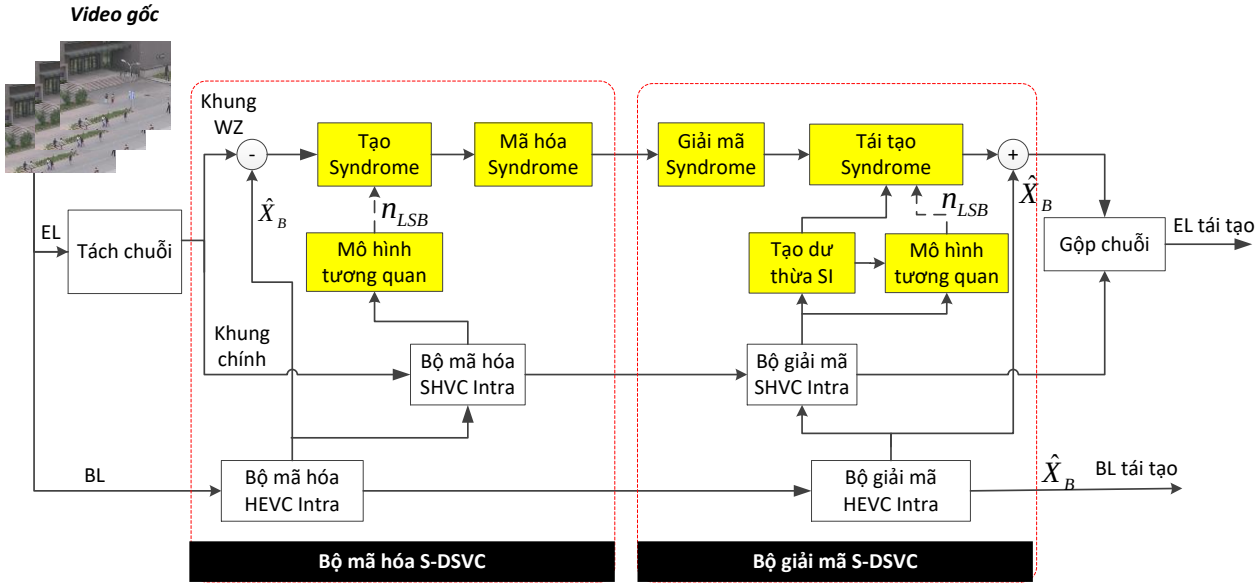
- Kết hợp khung: Các khung hình WZ và khung hình chính được hợp nhất với nhau tạo nên lớp nâng cao.

4.3. Đề xuất trên DSVC

Thành tựu nghiên cứu hiện tại của các tiêu chuẩn mã hóa video khả chuyển khá hạn chế, đáng chú ý là trong các ứng dụng đa phương tiện mới nổi như mạng cảm biến hình ảnh, hệ thống giám sát video và truyền không gian từ xa, thường yêu cầu tiêu thụ năng lượng thấp phía mã hóa video để sử dụng trong thời gian dài và khả năng chịu lỗi cao, nhận video liền mạch. Để giải quyết vấn đề này, luận văn này đề xuất một giải pháp mã hóa video khả chuyển phân tán (S-DSVC) thế hệ mới, kết hợp tiêu chuẩn H.265/HEVC gần đây nhất với mô hình mã hóa video phân tán (DVC) đầy hứa hẹn. S-DSVC được đề xuất nhằm mục đích sử dụng hiệu quả mã hóa của cả H.265/HEVC và DVC để đáp ứng mức tiêu thụ năng lượng thấp và khả năng chịu lỗi cao.

4.3.1. Kiến trúc S-DSVC

Hình 4.5 minh họa kiến trúc của giải pháp mã hóa video giám sát được đề xuất, trong đó các thành phần mã hóa phân tán mới được làm nổi bật. Cách tiếp cận được đề xuất cũng tuân theo cách tiếp cận mã hóa liên lớp để cung cấp tính năng mở rộng. Khái niệm mã hóa phân tán được sử dụng ở các lớp nâng cao trong khi mô hình mã hóa video dự đoán, đáng chú ý là HEVC được sử dụng ở lớp cơ bản. Để đạt được yêu cầu độ phức tạp tính toán thấp, cả hai lớp cơ bản và lớp nâng cao đều được mã hóa Intra; do đó, dẫn đến độ phức tạp tính toán thấp ở phía bộ mã hóa.



Hình 4.5. Cấu trúc S-DSVC

Ý tưởng cơ bản của giải pháp được đề xuất là phần EL dư thừa được mã hóa khai thác một số tương quan thời gian theo cách phân tán [20], và do đó chỉ là một phần của EL dư thừa không thể ước tính được bằng cách tạo thông tin phụ trợ bên bộ giải mã (SI), được mã hóa và gửi đến bộ giải mã. Để tránh gửi thông tin có thể được suy ra tại bộ giải mã, mô hình tương quan (Correlation model - CM) xác định số lượng bit ít quan trọng nhất, phần khác nhau giữa EL dư thừa và SI được mã hóa và truyền đi.

Đối với mã hóa EL, cách tiếp cận DVC đã được sử dụng trong phương pháp được đề xuất, trong đó các khung video đầu vào được chia thành hai phần: khung chính và khung WZ như trong Hình 4.5. Trong phương pháp này, các khung chính được mã hóa bằng SHVC thông thường [6] trong khi các khung WZ được mã hóa bằng cách sử dụng tạo Syndrome, mã hóa Syndrome và mô hình tương quan. Tại bộ giải mã, dòng bit nhận được được xử lý để thu được dữ liệu video gốc bằng cách giải mã Syndrome, tái tạo Syndrome, mô hình tương quan và tạo thông tin phụ trợ (SI). Trong sơ đồ mã hóa như vậy, các tính năng phức tạp thấp của DVC một lần nữa được khai thác hiệu quả trong phương pháp này trong đó cả khung chính và khung WZ được mã hóa bằng cách sử dụng phương pháp mã hóa Intra đơn giản và truyền; do đó, không có ước tính chuyển động phức tạp nào được thực hiện tại bộ mã hóa S-DSVC được đề xuất [24].

Tóm lại, chuỗi các bước mã hóa EL có thể được tóm tắt là:

E1) Tách chuỗi: Đầu tiên, các khung EL được chia thành các khung chính và WZ. Số lượng khung WZ giữa hai khung chính liên tiếp được xác định bởi kích thước GOP. Đương nhiên, kích thước GOP là 2 thường được sử dụng do sự cân bằng giữa hiệu quả nén và yêu cầu độ trễ giải mã của video.

E2) Tạo Syndrome: Đối với các khung WZ, phần thông tin EL dư được tạo bằng cách trừ khung giải mã BL từ khung gốc. Phần dư này sau đó được biến đổi với biến đổi côsin rời rạc số nguyên (DCT) và được lượng tử vô hướng với kích thước bước lượng tử hóa EL để tạo ra dư lượng lượng tử EL. Trong giải pháp S-DSVC được đề xuất, chỉ một phần dư lượng tử EL, được gọi là Syndrome, được mã hóa và gửi đến bộ giải mã. Kích thước Syndrome được đặc trưng chủ yếu bởi sự tương quan giữa dư lượng ban đầu và dư lượng SI được tạo ra tại bộ giải mã.

E3) Mô hình tương quan (CM): Để nén hiệu quả dư lượng EL, mối tương quan giữa dư lượng EL gốc và dư lượng SI phía giải mã được ước tính ở bước này. Ở đây, mức độ tương quan được xác định thông qua một số bit có ý nghĩa nhỏ nhất, n_{LSB}, cần được truyền đến người nhận. Trong đề xuất này, n_{LSB} có thể được tính tương tự như được trình bày trong [28].

E4) Mã hóa Syndrome: Syndrome được tạo ra từ bước trước sẽ được nén bằng cách sử dụng giải pháp mã Context Adaptive Binary Arithmetic Coding (CABAC) như trong các tiêu chuẩn mã hóa video dự đoán như H.264/AVC và HEVC.

Tại phía thu, chuỗi các bước giải mã EL bao gồm:

D1) Giải mã Syndrome: Đầu tiên, Syndrome EL nhận được được giải mã bằng cách sử dụng giải pháp giải mã Context Adaptive Binary Arithmetic Decoding (CABAD). Syndrome là một phần quan trọng của thông tin ban đầu không thể ước tính được ở bộ giải mã sử dụng giải pháp tạo thông tin phụ trợ (SI) được trình bày trong bước tiếp theo.

D2) Tạo dư thừa SI: SI là phiên bản nhiễu của thông tin gốc có thể được tạo ở phía bộ giải mã. Đương nhiên, chất lượng SI cao hơn, bitrate thấp hơn là cần thiết để gửi đến bộ giải mã. Do đó, chất lượng SI đóng vai trò cực kỳ quan trọng trong giải pháp S-DSVC được đề xuất. Xem xét mối tương quan thời gian cao giữa các khung liên tiếp trong chuỗi video giám sát, trong bài viết này đề xuất một giải pháp tạo SI hiệu quả như được mô tả trong mục nhỏ tiếp theo.

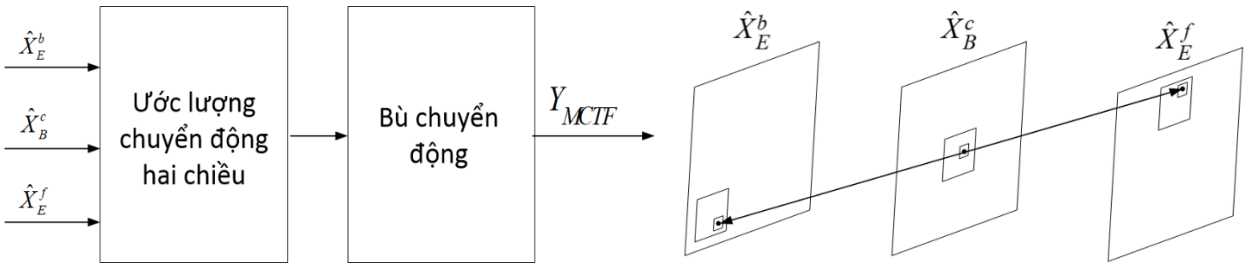
D3) Mô hình hóa tương quan: Tương tự như bộ mã hóa, mô hình tương quan được tiến hành trong bộ giải mã cũng nhằm mục đích ước tính mối tương quan giữa khung

gốc và dư lượng SI. Mối tương quan này cũng được thể hiện thông qua một số lượng bit có ý nghĩa và được tính toán như ở phía bộ mã hóa.

D4) Tái tạo Syndrome: Cuối cùng, thông tin EL được xây dựng lại bằng cách sử dụng Syndrome được gửi từ người nhận và dư lượng SI được tính toán tại bộ giải mã. Để đạt được khung EL chất lượng cao nhất, một giải pháp tái cấu trúc thống kê như được trình bày trong [29] được áp dụng.

4.3.2. Đề xuất tạo khung SI

Để tạo khung SI, luận văn đề xuất một sơ đồ mới, cụ thể là Lọc thời gian bù chuyển động (Motion Compensated Temporal Filtering - MCTF), có thể khai thác hiệu quả tính tương quan thời gian cao giữa hai khung chính EL liên tiếp được đặc trưng cho video giám sát. Hình 4.6 cho thấy sơ đồ MCTF được đề xuất trong đó các khung đầu vào bao gồm khung BL hiện tại đã giải mã, khung giải mã EL trước và sau, $\hat{X}_B^c$, $\hat{X}_E^f$, $\hat{X}_E^b$, tương ứng.



a) Sơ đồ MCTF b) Ước lượng chuyển động 2 chiều

Hình 4.6. Sơ đồ MCTF đề xuất

Như được trình bày trong Hình 4.6, mối tương quan thời gian được khai thác để cải thiện chất lượng khung BL bằng cách tìm sự dịch chuyển của từng khối BL (chất lượng thấp hơn) trong hai khung EL (chất lượng cao hơn), sau đó lấy trung bình các khối EL dịch chuyển và khối BL để đạt được khung SI cuối cùng. Do đó, MCTF có thể được thực hiện như sau:

- *Ước lượng chuyển động hai chiều (Bi-directional Motion Estimation - BiME*): Bước này nhằm tìm một tập hợp các vecto chuyển động (Motion Vector – MV) thể hiện tốt chuyển động của từng khối BL được giải mã $\hat{X}_B^c$ đối với khung EL phía sau đã giải mã, $\hat{X}_E^b$ và khung EL phía trước đã giải mã, $\hat{X}_E^f$. BiME sẽ tạo ra một cặp MV đối xứng, một MV chỉ vào $\hat{X}_E^b$ và một MV khác chỉ vào $\hat{X}_E^f$.

- *Bù chuyển động (Motion compensation - MC):* Sử dụng các MV thu được từ bước trước, hai ứng cử viên ước lượng SI ($\hat{X}_E^f(i+\overrightarrow{mv})$, $\hat{X}_E^b(i-\overrightarrow{mv})$) bằng cách thực hiện bù chuyển động dựa trên hai khung tham chiếu EL trước và sau. Tiếp theo, trung bình ước lượng bù chuyển động này và khung BL được tính để thu được khung MCTF SI, Y_{MCTF}, như sau:

$$Y_{MCTF}(i) = \frac{1}{3}[\hat{X}_E^b(i-\overrightarrow{mv}) + \hat{X}_B^c(i) + \hat{X}_E^f(i+\overrightarrow{mv})] \quad (8)$$

trong đó i là khối thứ i trong khung.

Như được xác định trong (8), khung MCTF SI, Y_{MCTF}, được tạo ra không chỉ từ BL đã giải mã mà còn từ bù chuyển động hai khung có nguồn gốc từ các khung được giải mã EL phía trước và phía sau để xem xét cả tương quan không gian và thời gian. Điều này có thể đảm bảo chất lượng SI tốt ngay cả khi khung giải mã BL có chất lượng thấp hơn.

4.4. Kết quả

4.4.1. Các điều kiện thử nghiệm

Việc đánh giá hiệu năng cho giải pháp S-DSVC được thực hiện đối với sáu video giám sát thu được từ bộ dữ liệu PKU-SVD-A [12, 13]. Hình 4.7 cho thấy các khung đầu tiên của các video giám sát được thử nghiệm trong khi Bảng 3 tóm tắt một số đặc điểm chính của chúng và các tham số lượng tử hóa được sử dụng để nén khung BL và EL. Như thường lệ, kết quả được trình bày cho thành phần độ chói và với tất cả các khung (khung BL, khung chính EL và khung WZ EL).

Hình 4.7. Khung đầu tiên của các video thử nghiệm

Bảng 3. Điều kiện thử nghiệm S-DSVC

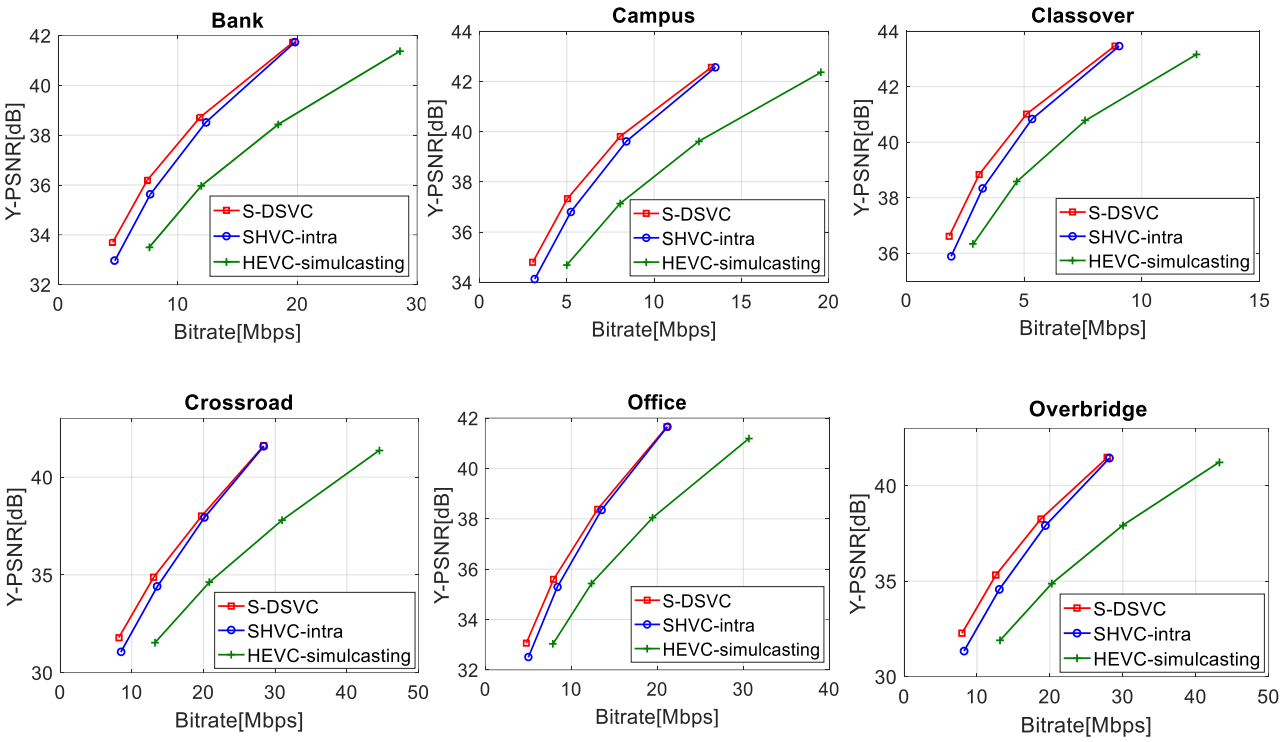
Kích thước, tốc độ khung hình, số khung được nén	720×576, @30Hz, 201 frames
Kích thước GOP	2 (Key-WZ-Key-…)
Thông số lượng tử	QP_B = {38;34;30;26} $QP_E = QP_B - 4$

4.4.2. Đánh giá hiệu năng

Như đã đề cập ở trên, trong nghiên cứu mã hóa video, RD performance thường được sử dụng để đánh giá một giải pháp mã hóa video mới. Trong bối cảnh này, hai chuẩn mã hóa video giám sát phù hợp nhất được so sánh với giải pháp S-DSVC được đề xuất, đáng chú ý là giải pháp SHVC-intra [6] và giải pháp HEVC-simulcasting. Cần lưu ý rằng, SHVC-intra được thực hiện bằng cách nén dữ liệu video giám sát bằng phần mềm SHVC [30] và cấu hình mã hóa Intra trong khi HEVC-simulcasting được thực hiện bằng cách nén dữ liệu video giám sát bằng phần mềm HEVC [31] và với hai lớp độc lập. So sánh RD performance được hiển thị trong Hình 4.8 trong khi Bảng 4 thể hiện mức tiết kiệm BD-Rate [32] khi so sánh giải pháp đề xuất S-DSVC với các chuẩn liên quan.

Bảng 4. Tiết kiệm BD-Rate S-DSVC

Chuỗi video	**SHVC-intra vs. HEVC-simulcasting**	**S-DSVC vs. HEVC-simulcasting**	**S-DSVC vs. SHVC-intra**
Bank	-32,85	-39,19	-9,04
Campus	-28,93	-38,54	-9,41
Classover	-28,93	-36,83	**-10,58**
Crossroad	**-34,58**	-38,62	-5,93
Office	-32,41	-37,08	-6,55
Overbridge	-34,14	**-40,56**	-9,46
Trung bình	***-31,97***	***-38,47***	***-8,49***



Hình 4.8. So sánh RD performance S-DSVC

Từ kết quả thu được, có thể rút ra một số kết luận:

- Như được hiển thị trong Hình 4.8 và Bảng 4, giải pháp S-DSVC được đề xuất vượt trội đáng kể so với tiêu chuẩn HEVC-simulcasting với mức tiết kiệm bitrate khoảng 38,5% trong khi vẫn duy trì chất lượng tương tự.
- Đạt được hiệu năng nén cao hơn cho các chuỗi video chứa ít chuyển động hơn, ví dụ: *Overbridge* và *Classover*. Điều này là do việc tạo thông tin phụ trợ bên trong cấu trúc mã hóa được đề xuất thường hoạt động tốt đối với nội dung video chuyển động thấp, do đó mang lại hiệu quả nén cao cho giải pháp S-DSVC được đề xuất.
- S-DSVC được đề xuất cũng đạt được mức tăng nén tốt hơn khi so sánh với tiêu chuẩn SHVC thông thường, đáng chú ý là với mức tiết kiệm bitrate khoảng 8,5%.

4.4.3. Đánh giá độ phức tạp S-DSVC

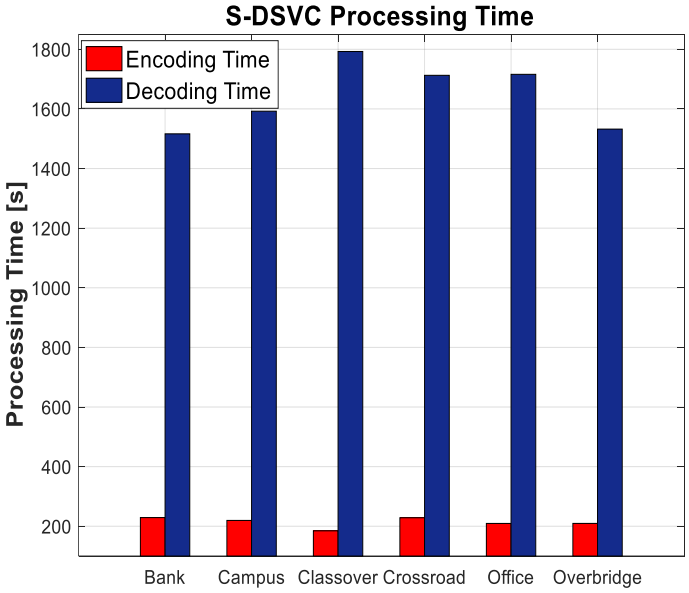
Trong phần này, chúng tôi đánh giá sự phức tạp liên quan đến kiến trúc S-DSVC được đề xuất cũng như so sánh với tiêu chuẩn SHVC nổi tiếng. Để đạt được đối tượng này, thời gian xử lý [giây] được sử dụng như một đại diện cho độ phức tạp tính toán của từng giải pháp mã hóa. Cấu hình của máy tính được sử dụng để thử nghiệm được chỉ định trong Bảng 5.

Bảng 5. Cấu hình phần cứng để thử nghiệm S-DSVC

Hardware configuration	• **Processor:** Intel® Core™ i7-4800MQ @2.7 GHz • **RAM:** 8.00 GB • **System:** Win 10, 64-bit • **Environment:** Microsoft Visual Studio 2017 Community

4.4.3.1. Phân tích thành phần S-DSVC

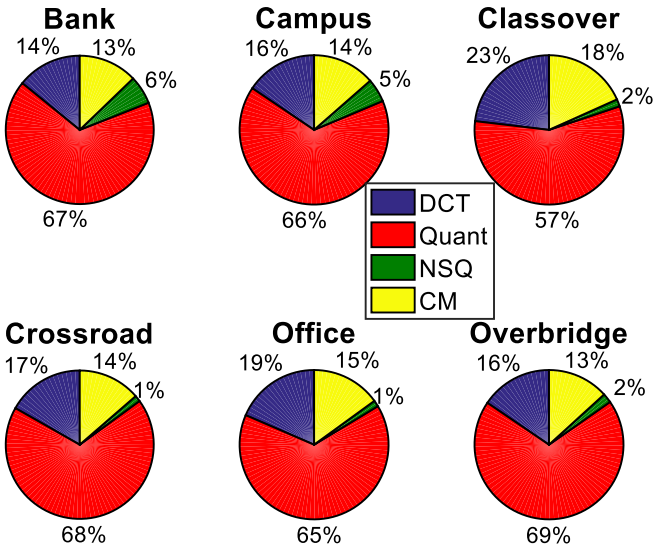
Như đã thảo luận, trái ngược với các tiêu chuẩn mã hóa video dự đoán thông thường [3, 4], S-DSVC được đề xuất chuyển một trong những phần phức tạp tính toán nhất, ước tính chuyển động, sang phía bộ giải mã. Điều này dẫn đến một giải pháp mã hóa video phức tạp thấp. Để hiểu được hiệu ứng này, luận văn đã đo lường và so sánh độ phức tạp liên quan đến từng mặt mã hóa (bộ mã hóa và bộ giải mã) cho codec S-DSVC được đề xuất. Hình 4.9 cho thấy sự so sánh giữa các quá trình mã hóa và giải mã cho sáu video giám sát được thử nghiệm.



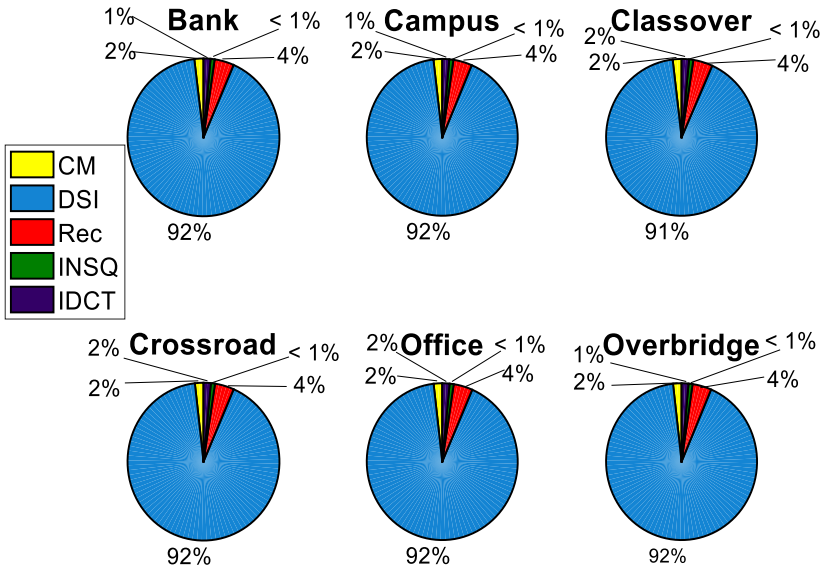
Hình 4.9. So sánh thời gian mã hóa và giải mã

Như được hiển thị, độ phức tạp tính toán liên quan đến phía bộ mã hóa thấp hơn nhiều so với phía bộ giải mã. Để hiểu rõ hơn về độ phức tạp tính toán liên quan đến từng

công cụ mã hóa của giải pháp S-DSVC được đề xuất, phân tích độ phức tạp cho từng thành phần được thực hiện cho cả bộ mã hóa và bộ giải mã và được hiển thị trong Hình 4.10 và Hình 4.11, tương ứng.



Hình 4.10. Thời gian mã hóa – các thành phần



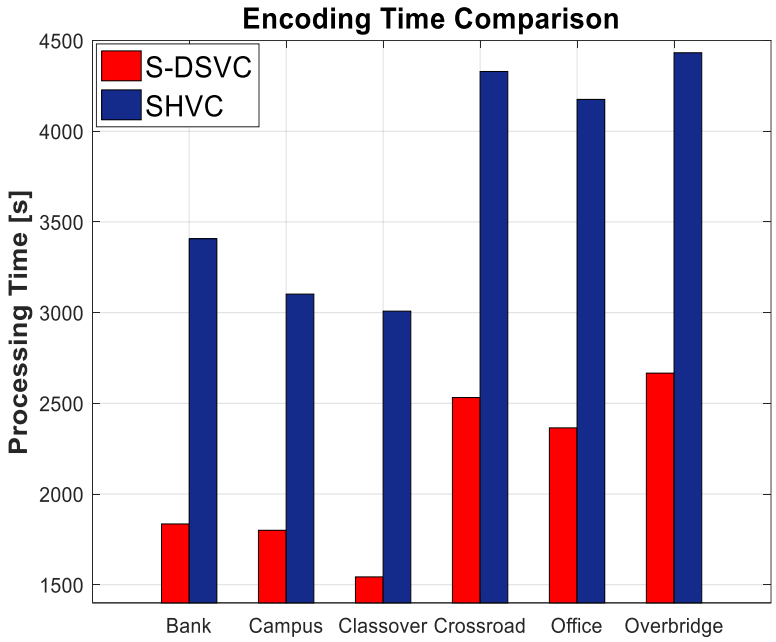
Hình 4.11. Thời gian giải mã – Các thành phần

Có thể thấy rằng trong quá trình mã hóa, Lượng tử hóa (Quant) tiêu thụ phần trăm thời gian xử lý cao nhất, khoảng 60-70% và tiếp theo là Biến đổi Cosine rời rạc (DCT), Mô hình tương quan (CM) và Lượng tử vô hướng lồng nhau (NSQ).

Ở phía bộ giải mã, Thông tin phụ trợ phía bộ giải mã (DSI) tiêu thụ tỷ lệ phần trăm xử lý cao nhất, với khoảng 90%. Điều này chủ yếu là do sự phức tạp cao - quá trình ước tính chuyển động của DSI. Các thành phần khác, mô hình tương quan (CM), nghịch đảo DCT (IDCT), tái cấu trúc (Rec) và ngịch đảo lượng tử vô hướng lồng nhau (INSQ) tiêu thụ ít hơn 10% thời gian giải mã tổng thể.

4.4.3.2. S-DSVC so với SHVC

Một trong những lợi ích chính của giải pháp S-DSVC là độ phức tạp tính toán liên quan đến phần mã hóa. Để chứng minh sự tiến bộ này, chúng tôi so sánh thời gian mã hóa [giây] của S-DSVC được đề xuất với điểm chuẩn SHVC. Thí nghiệm này được thực hiện trong sáu chuỗi video thử nghiệm và được hiển thị trong Hình 4.12.



Hình 4.12. So sánh thời gian mã hóa

Như có thể thấy trong Hình 4.12, độ phức tạp liên quan đến giải pháp S-DSVC được đề xuất thấp hơn nhiều so với tiêu chuẩn SHVC, đáng chú ý là giảm khoảng 60% thời gian mã hóa. Tính năng quan trọng này làm cho giải pháp S-DSVC được đề xuất phù hợp với một số lượng lớn các ứng dụng giám sát video, thường bị hạn chế bởi nguồn và năng lượng.

KẾT LUẬN

Luận văn này trình bày các giải pháp mã hóa video khả chuyển để nén nội dung hình ảnh cho hệ thống giám sát. Một giải pháp mã hóa video khả chuyển cho hệ thống giám sát được xây dựng trên nền chuẩn SHVC và khai thác đặc tính chuyển động thấp của video giám sát thông qua cơ chế lựa chọn tham chiếu dài hạn thích ứng. Theo đánh giá, SSVC được đề xuất vượt trội đáng kể so với tiêu chuẩn SHVC. Trong tương lai có thể xem xét để cải thiện hơn nữa tính chính xác của cơ chế lựa chọn tham chiếu dài hạn hoặc tạo ra một giải pháp mã hóa video giám sát tốt hơn để tính đến chất lượng của khung tham khảo dài hạn cho giải pháp SSVC.

Ngoài ra giải pháp S-DSVC theo cách tiếp cận mã hóa theo lớp trong khi kết hợp hiệu quả giữa mã hóa video dự đoán và phân tán cũng mang lại hiệu quả cao khi so với các chuẩn mã hóa liên quan. Phương pháp mã hóa phân tán được sử dụng để nén dữ liệu lớp nâng cao trong khi tiêu chuẩn HEVC được sử dụng để nén dữ liệu của lớp cơ bản. Giải pháp được chọn này có thể khai thác mối tương quan thời gian giữa các khung video giám sát tại bộ giải mã đồng thời đảm bảo khả năng tương thích ngược với HEVC nổi tiếng ở lớp cơ bản.

TÀI LIỆU THAM KHẢO

[1] Cisco Visual Networking Index: Forecast and Methodology, 2015 - 2020, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html

[2] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[3] P. N. Tudor, "MPEG-2 Video Compression," *Electronics and Communication Engineering Journal*, pp. 257 – 264, Dec. 1995.

[4] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Circuits and Systems for Video Tecnology*, vol. 13, no. 7, pp. 560-576, Jul. 2003.

[5] H. Schwarz, D. Marpe, and T. Wiegand "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103-1120, Sep. 2007.

[6] ISO/IEC JTC 1/SC 29/WG 11 and ITU-T SG16 WP3, "Joint call for proposals on scalable video coding extensions of High Efficiency Video Coding (HEVC)," Doc. N12957, Stockholm, Sweden, Jul. 2012.

[7] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 301-317, March 2001.

[8] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile Video Transmission using Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1204-1217, vol. 17, no. 9, Sep. 2007.

[9] P. Baccichet, T. Schierl, T. Wiegand, and B. Girod, "Low -delay Peer-to-Peer Streaming," *Packet Video Workshop*, pp. 173-181, Lausanne, Switzeland, Nov. 2007.

[10] P. L. Dragotti and M. Gastpar (editors), "Distributed Source Coding: Theory, Algorithms and Applications," *Academic Press*, 2009.

[11] M. Valera and S. Velastin, "Intelligent distributed surveillance systems: A review," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 152, no. 2, pp. 192–204, Apr. 2005.

[12] W. Gao, Y. Tian, T. Huang, S. Ma, and X. Zhang, "IEEE 1857 standard empowering smart video surveillance systems," *IEEE Intelligent Systems*, 2013.

[13] PKU-SVD-A. [Online]. Available: http://mlg.idm.pku.edu.cn/-resources/pku-svd-a.html

[14] "Video test sequences," [Online]. Available: ftp://hevc@ftp.tnt.uni-hannover.de/testsequences/

[15] X. Zhang, L. Liang, Q. Huang, T. Huang, W. Gao, "A background model based method for transcoding surveillance videos captured by stationary camera," *IEEE Picture Coding Symposium (PCS),* Nagoya, Japan, pp. 78-81, 2010.

[16] X. Zhang, T. Huang, Y. Tian, and W. Gao, "Background-modeling-based adaptive prediction for surveillance video coding," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 769–784, 2014.

[17] X.G. Zhang, L.H. Liang, Q. Huang, Y.Z. Liu, T.J. Huang, and W. Gao, "An efficient coding scheme for surveillance videos captured by stationary cameras," *IEEE International Conference on Visual Communication and Image Processing (VCIP),* pp. 77442A1–10, 2010.

[18] S. Zhang, K. Wei, H. Jia, X. Xie, W. Gao, "An efficient foreground-based surveillance video coding scheme in low bit-rate compression," *IEEE International Conference on Visual Communication and Image Processing (VCIP),* San Jose, USA, Nov. 2012.

[19] X. Zhang, Y. Tian, T. Huang, S. Dong, W. Gao, "Optimizing the Hierarchical Prediction and Coding in HEVC for Surveillance and Conference Videos with Background Modeling," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4511-4526, Oct. 2014.

[20] F. Pereira, L. Torres, C. Guillemot, T. Ebrahimi, R. Leonardi, and S. Klomp, "Distributed video coding: selecting the most promising application scenarios," *Signal Processing: Image Communication*, vol. 23, no. 5, pp. 339–352, June 2008.

[21] D. Slepian, J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Information Theory Society*, vol. 19, pp. 471–480, 1973.

[22] A.D. Wyner, J. Ziv, "The rate-distortion function for source coding with side information at the decoder*," IEEE Information Theory Society*, vol. 22, no. 1, pp. 1–10, 1976.

[23] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, M. Ouaret, "The DISCOVER Codec: Architecture, Techniques and Evaluation," *IEEE Picture Coding Symposium (PCS),* Lisboa, Portugal, Nov. 2007.

[24] R. Puri, A. Majumdar, and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2436-2448, Oct. 2007.

[25] F. Pereira, C. Brites, J. Ascenso, M. Tagliasacchi "Wyner–Ziv video coding: a review of the early architectures and further developments," *IEEE International Conference on Multimedia and Expo (ICME),* Hannover, Germany, June 2008.

[26] L. Liu, Z. Li, and E. J. Delp, "Efficient and Low-Complexity Surveillance Video Compression Using BackwardChannel Aware Compression," *IEEE Circuits and Systems for Video Technology,* vol. 19, no. 4, Apr. 2009.

[27] V. K. Kolavella and P. G. Krishna Mohan "Distributed video coding: codec architecture and implementation," *Signal and Image Processing: An International Journal (SIPIJ),* vol. 2, no. 1, pp. 151-163, Mar. 2011.

[28] X. Hoang Van, J. Ascenso, F. Pereira, "HEVC backward compatible scalability: A low encoding complexity distributed video coding based approach," *Signal Processing: Image Communication*, vol. 33 pp. 51–70, Apr. 2015.

[29] X. HoangVan, J. Ascenso, and F. Pereira, “Optimal Reconstruction for a HEVC Backward Compatible Distributed Scalable Video Codec,” *IEEE Visual Communication and Image Processing (VCIP),* Valletta, Malta, Dec. 2014.

[30] SHVC reference software, [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSoftware/.

[31] HEVC reference software, [Online]. Available: https://hevc.hhi.fraun hofer.de/svn/svn_HEVCSoftware/.

[32] G. Bjontegaard, “Improvements of the BD-PSNR model,” *document ITU-T SC16/Q6, Doc. VCEG-AI11,* 2008.