

일반논문 (Regular Paper)

방송공학회논문지 제29권 제6호, 2024년 11월 (JBE Vol.29, No.6, November 2024)

<https://doi.org/10.5909/JBE.2024.29.6.999>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

임의접근모드의 목표비트율 정확도 향상을 위한 비디오 비트율-제어

현명한^{a)‡}

A Video Bit-Rate Control for Enhancing Target Bit-Rate Accuracy in VVC Random Access Mode

Myung Han Hyun^{a)‡}

요약

비디오 영상 비트율-제어 기술은 효율적인 영상전송에 필수적인 기술이다. 고해상도 및 다양한 용도의 영상에 대한 수요가 늘어남에 따라 한정된 대역폭 또는 저장공간에 전송·저장이 가능하도록 비트율을 조절해야 한다. 기존의 R- λ 모델을 이용한 비디오 비트율-제어 기법은 이미 상당한 수준으로 목표비트율을 맞추고 있다. 하지만 임의접근 (random access, RA) 모드를 이용한 영상 부호화에서는 여전히 목표비트율을 맞추기가 어렵다. 따라서 목표비트율을 효율적으로 달성하기 위해 비트율-왜곡 최적화 (rate distortion optimization, RDO)에 사용되는 λ 값을 조정하여 비트를 더욱 효과적으로 할당할 필요가 있다. 본 논문에서는 효율적인 λ 값 조정을 통한 임의접근모드에서의 최적 목표비트율 제어방법을 제안한다. 목표비트율 정확도 평가기준으로 성능을 평가한 결과, 제안방법이 기존 방법 대비 평균 7.82% 포인트 비트율정확도가 향상됨을 확인할 수 있었다.

Abstract

Video bit-rate control is essential for efficient video transmission. As the demand for high-resolution and versatile video applications increases, adjusting the bit rate to enable transmission/storage within a limited bandwidth or storage device is necessary. The traditional video bit-rate control method using the R- λ model achieves a higher target bit-rate accuracy. However, it remains challenging to achieve the target bit-rate accuracy in random access (RA) mode. Thus, it is necessary to adjust the λ value used in rate-distortion optimization (RDO) to allocate bits more effectively. Therefore, this paper proposes an optimal target bit-rate control method in RA mode through efficient λ value adjustment. Based on the target bit-rate accuracy criteria, the proposed method achieved an average bit-rate accuracy improvement of 7.82% point compared to the previous method.

Keyword : Video Bit-Rate Control, Versatile Video Coding, Random Access, Bit-Rate Accuracy, R- λ model

a) 경상국립대학교 IT공과대학 제어로봇공학과 및 공학연구원(Department of Control and Robot Engineering, College of IT Engineering and Engineering Research Institute (ERI), Gyeongsang National University (GNU))

‡ Corresponding Author : 현명한(Myung Han Hyun)

E-mail: drwise@gnu.ac.kr

Tel: +82-55-772-1743

ORCID: <https://orcid.org/0000-0002-8364-6127>

· Manuscript August 21, 2024; Revised October 9, 2024; Accepted October 10, 2024.

I. 서론

비디오 영상 부호화는 초창기 화상회의용 비디오 압축 기술로부터 시작하여 디지털 TV 방송, 케이블 TV, 디지털 카메라, 스마트폰, 영상 스트리밍 등등의 다양한 멀티미디어 애플리케이션에 사용되는 기술로서 현재까지 지속적으로 발전하고 있다. 또한 스마트폰과 같은 다양한 영상 촬영 및 제작 기기의 보급으로 말미암아 생성된 방대한 영상 콘텐츠는 멀티미디어 처리 기술 연구를 가속화 시키고 있다. 최근 표준화가 완료된 Versatile Video Coding (VVC)은 이전 영상압축 표준인 High Efficient Video Coding (HEVC) 보다 더 큰 해상도의 영상 처리에 적합한 표준으로서, 해상도 적응 비디오 전송 (adaptive resolution video streaming), 초저지연 비디오 전송 (ultra low-delay streaming), 스크린 콘텐츠 비디오 부호화 (screen content video coding), 전방위 몰입영상 부호화 (omnidirectional immersive video coding), 광색역 비디오 부호화 (wide color gamut video coding) 등과 같은 용도로 사용될 수 있다^[1-2].

비디오 영상 부호화의 핵심 기술인 비트율-제어 (rate control, RC) 기법은 목표 화질 또는 목표비트에 맞게 부호화 비트를 생성하여 실제 통신 환경에서 효율적인 영상 스트리밍을 가능케 하는 영상압축 기술이다. 따라서 비디오 영상물 편집 또는 보관, 디지털 카메라의 비트율-제어 영상압축 등 제한된 환경에서의 저장 및 스트리밍에 비트율-제어가 필수적이다. 비록 비트율-제어 관련 기술은 표준화 대상 기술은 아니지만, 비디오 영상 부호화 기술 연구가 시작된 시점부터 지속적인 연구가 수행되고 있으며 실용성이 대단히 높은 기술이다. 비트율-제어를 가장 손쉽게 하는 방법은 부호화되는 프레임의 할당 비트를 기계적으로 할당하되 목표비트를 남은 프레임 수로 나누어 할당하는 방법이다. 이 방법은 비트율-왜곡 최적화 (rate-distortion optimization, RDO) 관점에서 효율이 낮기 때문에, 같은 비트를 사용하되 최소의 화질왜곡을 가지는 알고리즘이 더 성능이 좋다고 평가할 수 있다. 그러나 비트율-제어는 각각의 애플리케이션 마다의 특성을 고려한 다양한 방법으로 비트율-제어가 이루어지기 때문에 해당 특성에 적합한 방법으로도 평가될 수 있다. 다시 말하면 화질왜곡 측면에서 급격한 화질의 변동을 줄이는 방향으로 비트율-제어 평가가 이

루어질 수도 있으며, 통신 대역의 효율성을 높일 수 있는 비트 항상성에 방점을 두는 방향으로도 비트율-제어 평가가 이루어질 수 있다. 특히 비트 항상성에 중점을 둔 비트율-제어 기술은 국방전술데이터 영상전송과 같은 통신 대역의 제약이 큰 분야의 대역 효율성을 높이는 데 큰 기여를 할 수 있다.

비트율-제어의 근간이 되는 기술이 비트율-왜곡 (rate-distortion, R-D) 모델링 기술이다. VVC 표준화가 진행되는 중에 VVC 표준 부호화 구조에 적합한 R-D 모델링 기법이 연구되어 비트율 및 화질왜곡 예측 성능을 크게 향상시켰다^[3]. 또한 이 방법은 Recursive Bayesian Estimation (RBE) 기반 비트율-제어 방법으로 확장되어 화질의 항상성을 높이는 비트율-제어 기술로 발전하였다^[4]. 아울러 Lagrange Multiplier Adjustment (LMA) 기반 비트율-제어 알고리즘이 연구되었고 이는 통신 제약이 크면서 영상데이터의 중요도가 큰 애플리케이션의 비트 항상성을 높이는 방법으로 화면내 예측 (all intra, AI) 모드를 비트율-제어에 적용하여 효율성을 검증하였다^[5]. 하지만 임의접근 (random access, RA) 모드에서의 비트율-제어에서는 화면그룹 (group of pictures, GOP)마다 주기적으로 삽입되는 화면내 (intra, I) 프레임 때문에 최종 부호화된 비트를 처음 설정한 목표비트에 맞게 생성하기가 쉽지 않다. 왜냐하면, I-프레임은 B (bi-directionally) 프레임, P (predictive-coded) 프레임 대비 많은 비트가 사용되어 부호화되기 때문에 비트율 정확도를 맞추기가 어렵고 RA 모드의 부호화도 계층적 구조로 진행되어 비트율-제어도 다른 예측모드에 비해 다소 복잡하기 때문이다. 즉 최종 부호화 비트와 설정된 목표비트 간의 불일치성은 통신 대역 자원의 낭비를 초래하고 부호화 효율성을 떨어뜨린다. 따라서 본 논문에서는 RA 모드 비트율-제어 사용 시에 목표비트 정밀도를 높이기 위한 방법으로서 RDO에 사용되는 λ 값을 추정 및 적용하는 방법에 대해 제안한다. 본 논문에서 제안하는 기본 알고리즘은 기존 AI 모드에서의 비트율-제어에 적용된 기법을 바탕으로 제안된 것이며^[6], 계층적 부호화 구조에 맞게 변형 적용되었다.

본 논문의 제 II장에서는 VVC 표준의 부호화 구조에 특화된 R-D 모델링 및 RC 기법의 관련 연구에 대해 소개하고, 제 III장에서 RA 모드의 계층적 부호화 구조에 맞게 변

형 적용된 비트율-제어 λ 값 추정 방법에 대해 서술한다. 제 IV장에서는 다양한 실험환경에서 수행한 실험결과에 대한 분석·평가·토의·고찰을 수행하고, 제 V장에서 결론을 맺는다.

II. 관련연구

비트율-제어 동작의 기본 원리는 부호화할 프레임의 비트를 미리 예측하여 그 비트를 적절히 생성해 내는 것에 있다. 이를 위해 양자화 파라미터 (quantization parameter, QP) 값을 계산하여 발생비트량을 조절하며 부호화를 진행함과 동시에 부호화에 사용할 수 있는 남아있는 비트를 나머지 프레임에 적절히 배분한다. 이러한 일련의 과정을 통하여 RDO 관점에서 부호화 효율을 높이고 동시에 목표비트 정밀도를 높이는 방법이 비트율-제어 기법이다. 이렇듯 비트율-제어 알고리즘이 부호화에 적용되면 비트율-제어 성능평가 측면에서 고려해야 될 사항이 많은데, 비트율-제어 성능향상을 위해 가장 기본이 되는 성능은 비트율 (R) 및 화질왜곡 (D) 수치 예측 성능이다. 이를 위해 R-D 모델링에 대한 연구가 필수적이며 정확한 R-D 모델을 바탕으로 정밀한 비트율-제어가 이루어지게 된다. 따라서 이어지는 소단원에서는 최신 R-D 모델링과 해당 모델을 적용한 비트율-제어 알고리즘에 대해 간략히 소개한다.

1. 파티클 필터링 기반의 비트율-화질왜곡 (R-D) 모델링

파티클 필터링은 유도항법제어, 로봇위치추정, 금융시장 분석과 같은 용도로 항공우주공학, 전자전기공학, 경제경영, 금융학 등의 다양한 분야에 활용되는 비선형 예측 기법이다^[6]. 비디오 영상에는 다채로운 색상 및 질감들이 분포하고 있으므로 비디오 영상 부호화 결과 나오는 비트나 화질의 값이 굉장히 무작위적이고 불안정한 특성이 있다. 최근의 연구에서 이러한 특성에 적합한 파티클 필터링을 이용한 R-D 모델이 제안되었고 큰 성능 향상을 나타냈다^[3]. 파티클 필터링은 기본적으로 무작위 값들로부터 하나의 대푯값을 예측해 내는 과정이라고 볼 수 있다. 그 예측한

대푯값이 실제값과 같다면 예측 오차는 '0'이 되고, 이상적인 예측이라고 할 수 있다. 그 과정의 첫 번째 과정인 '예측 부분'은 실제값이라고 추정되는 위치로 무작위 파티클들을 생성하는 부분이고, 둘째 과정인 '업데이트부분'은 실제 관측값을 바탕으로 무작위 파티클들의 확률을 업데이트하는 부분이다. 이러한 '예측부분'과 '업데이트부분'의 상호 보완과정을 통하여 얻어진 높은 확률의 파티클들은 정확한 실제값 예측에 중요한 역할을 담당하게 된다. 이렇게 획득된 예측 파티클들 값과 그것의 확률들을 내적하여 최종 추정값을 계산할 수 있다.

예측에 사용할 파티클을 생성하기 위해서는 가우시안 확률 분포를 갖는 임의의 파티클을 생성하고 그것들을 무작위로 추출하게 된다. 기존의 연구에서는 대략 150개 정도의 파티클을 사용하여도 R-D 예측성능이나 파티클 필터링의 계산 복잡도에서 큰 영향이 없음을 확인하였다^[3-5]. 생성된 파티클은 예측모델에 의해 예측하고자 하는 방향으로 전파되게 되며, 거기에서 관측값에 의해 각각의 파티클의 확률값이 재조정된다. 여기서 확률값이 재조정된 파티클들을 다시 한번 무작위 샘플링하고 각각의 파티클의 확률값을 정규화한다. 상기 파티클 필터링의 해당 과정 (예측, 확률값 재조정 및 정규화, 최종추정)은 다음의 소단원에서 수식으로 표현되었다.

기존 논문에서 제안하는 파티클 필터링 기반의 R-D 모델 생성을 위해서는 먼저 양자화 오차 및 엔트로피 계산식에 대한 정의가 필요하다^[3]. 그 계산식으로부터 R-D 모델 파라미터 값을 도출하는데에 파티클 필터링 기법이 적용된다. 이때 양자화 오차 및 엔트로피 계산식을 이용하려면 비디오 영상 부호화 양자화 후에 발생하는 변환계수의 확률 분포에 대한 정의가 필요하다. 이때의 확률 분포는 보통 라플라시안 확률 분포를 따른다고 알려져 있고 이 가정하에 R-D 모델은 양자화 오차 계산식과 엔트로피 이론에 근거하여 수식적으로 도출될 수 있다^[7]. 먼저 평균이 0인 라플라시안 확률 분포는 다음의 수식으로 표현될 수 있다.

$$l_x(x) = \frac{1}{2} \lambda \cdot e^{-\lambda|x|}, \lambda = \sqrt{2}/\sigma, x \in R \quad (1)$$

여기서 x 는 양자화 후 변환계수값이며, R 은 실수이고, σ

은 x 값의 표준편차이다. 여기서 라플라시안 파라미터 λ 는 다음과 같이 전개하여 계산할 수 있다.

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - E(x))^2 l_X(x) dx \\ &= \int_{-\infty}^{\infty} x^2 (\lambda/2) e^{-\lambda|x|} dx = 2/\lambda^2. \end{aligned} \tag{2}$$

다음의 수식은 화질왜곡 계산 수식이며, (2)에서 구한 라플라시안 파라미터와 q 값을 가지고 화질왜곡값을 계산할 수 있다. 여기서 q 는 양자화 계단 크기를 의미하며, δ 는 양자화 계산 때 사용되는 반올림 오프셋 값을 의미한다(단항 꼴의 수식은 참고문헌^[7]의 부록면 A-III를 참고).

$$\begin{aligned} D(\lambda, q) &= \int_{-\infty}^{\infty} (x - \hat{x})^2 l_X(x) dx \\ &= 2 \cdot \int_0^{(1-\delta)q} x^2 l_X(x) dx \\ &\quad + 2 \cdot \sum_{i=1}^{\infty} \int_{(i-\delta)q}^{(i+1-\delta)q} (x - iq)^2 l_X(x) dx \tag{3} \\ &= \lambda \cdot \int_0^{(1-\delta)q} x^2 e^{-\lambda x} dx \\ &\quad + \lambda \cdot \sum_{i=1}^{\infty} \int_{(i-\delta)q}^{(i+1-\delta)q} (x - iq)^2 e^{-\lambda x} dx \end{aligned}$$

또한 다음의 수식은 비트율 계산 수식이며, (2)에서 구한 라플라시안 파라미터와 q 값을 가지고 비트율 값을 계산할 수 있다. 여기에서 q 는 (3)의 것과 동일하다.

$$\begin{aligned} R(\lambda, q) &= - \sum_{i=-\infty}^{\infty} P_i \log_2 P_i = -P_0 \log_2 P_0 - 2 \sum_{i=1}^{\infty} P_i \log_2 P_i \\ P_0 &= \int_{-(1-\delta)q}^{(1-\delta)q} l_X(x) dx = 1 - e^{-(1-\delta)\lambda q} \tag{4} \\ P_i &= \int_{(i-\delta)q}^{(i+1-\delta)q} l_X(x) dx = \frac{e^{-(i-\delta)\lambda q} (1 - e^{-\lambda q})}{2}, \quad i > 0 \end{aligned}$$

파티클 필터링 기반의 R-D 모델은 (2)의 라플라시안 파라미터를 예측함에 있어서 이전 프레임의 부호화 결과인 평균제곱오차 (mean square error, MSE)로부터 임의의 화질왜곡 파티클을 생성하고 일련의 처리과정 (1. 파티클 예측생성, 2. 파티클의 확률값 재조정, 3. 파티클의 확률값 정규화, 4. 최종추정치 계산)을 거쳐 다음 프레임의 화질왜곡

값을 예측한다. 즉 파티클 필터링 기반의 R-D 모델링은 최종 예측된 화질왜곡값을 참값으로 가정하고 그 값을 부호화 하고자 하는 QP와 함께 (3)에 입력하여 라플라시안 파라미터를 추정하는 방법이다. 아울러 여기서 추정된 라플라시안 파라미터를 (4)에 적용하면 비트율 예측도 가능하다.

2. 파티클 필터링 기반의 비트율-제어

제 II장 1절에서 소개한 파티클 필터링 기반의 R-D 모델링 기법이 비트율-제어 응용에 확장되었고, 생성비트 배분을 효과적으로 하는데에 활용되어 화질의 항상성 성능을 높였다^[4]. 이는 부호화 이후의 화질왜곡값을 임의의 파티클로 생성하는 기존의 방법과는 달리, 부호화 이후의 부호화 비트값을 임의의 파티클로 적용·생성하여 미래 프레임의 비트율을 예측하는 파티클 필터링을 사용하였다. 이때 임의의 비트값 파티클을 예측·생성하면서 일정 프레임간의 화질왜곡의 비율을 사용하였고 그 비율은 다음과 같이 표현된다.

$$\Delta D_i = \alpha \cdot (D_i - D_{i-j}) / D_{i-j} \tag{5}$$

여기서 D 는 평균제곱오차를 의미하고, i 와 $(i-j)$ 는 프레임 번호를 의미한다. α 는 비트율 조정 파라미터로 실험에 사용된 모든 테스트 영상에 대해 0.3으로 고정하여 사용하였고, 이 수치가 비트율 변동의 예측을 효과적으로 수행함을 실험적으로 확인하였다.

(5)를 이용하여 생성된 다음 프레임의 비트값 파티클은 아래와 같이 표현된다.

$$\bar{b}_{i+1} = B_i \cdot \underbrace{[1 \ 1 \ 1 \ \dots \ 1]}_N^T + \Delta D_i \cdot \bar{b}_i \tag{6}$$

여기서 \bar{b}_{i+1} 및 \bar{b}_i 는 벡터로 표현된 $(i+1)$ 및 i 번째 프레임의 N 개의 비트값 파티클이며, B_i 는 비디오 영상 부호화 결과 얻어진 i 번째 프레임의 스칼라 비트값, ΔD_i 는 (5)의 수치이다. (6)의 비트값 파티클들은 무작위 재추출 및 확률정규화 과정을 거쳐 다음과 같이 표현된다(무작위 재

추출 방법은 참고문헌^[7]의 부록면 A-I을 참고).

$$c_{k,i+1} = p(b_{k,i+1}) / \sum_{l=1}^N p(b_{l,i+1}), k = 1, 2, \dots, N \quad (7)$$

여기서 $p(b_{k,i+1})$ 는 \bar{b}_{i+1} 의 확률밀도함수 (probability density function, pdf)를 의미하며, \bar{c}_{i+1} 는 확률정규화 과정을 거쳐 생성된 (i+1) 번째 프레임 파티클(N개)의 확률이다. 따라서 이렇게 생성된 (6)과 (7)의 비트값 파티클 및 해당 확률값의 내적으로 최종 예측 비트값이 계산되며 다음과 같이 표현된다.

$$\bar{B}_{i+1} = \bar{c}_{i+1} \cdot \bar{b}_{i+1} \quad (8)$$

(8)로 예측된 최종 예측 비트값은 비트율-제어에 사용되는 목표비트 할당에 적용되어 보다 정밀한 R-λ 모델^[8]의 λ 값을 예측한다. 결국 R-λ-QP의 관계식에 의해 새롭게 계산된 QP가 비디오 영상 부호화에 적용된다. 이러한 비트율 예측의 일련 과정을 거치면서 보다 화질왜곡 측면에서 화질변동이 적은 비트율-제어를 수행하였다.

III. 제안방법

제 II장에서의 관련연구 분석을 통하여 파티클 필터링이 R-D 모델링 및 비트율-제어에 효율적으로 응용되어 사용되고 있음을 확인하였다. 이러한 최신 연구의 비트율-제어 기법 적용에도 불구하고 생성비트의 정밀한 예측만으로는 비디오 영상 부호화의 비트변동률을 줄이기 어려운 실정이다.

특히 통신 대역이 협소하면서 중요 영상데이터의 손실이 없어야 하는 응용분야의 영상처리·전송에 큰 어려움이 있다. 열악한 통신 환경에서 중요데이터의 손실을 줄이기 위해서는 AI 모드의 사용이 필수적이다. 하지만 AI 모드의 사용으로 부호화 프레임마다의 비트 변동이 심하여 협소한 통신 대역 환경에 적합하지 않을뿐더러 비디오 영상 부호화 처리 결과를 임시 저장하는 버퍼 제어 측면에서도 비효율적이다. 따라서 상기 응용분야의 어려움을 해결하기 위해 RDO에 사용되는 λ값을 효율적으로 조절하여 비트변동을 줄이면서도 비트값의 항상성을 높이는 연구가 진행되었다^[5]. 앞서 서론에서도 언급하였지만 본 논문의 기본 알고리즘은 기존 AI 모드에서의 라그랑지 승수 조절을 통한 비트율-제어에 적용된 기법을 바탕으로 제안된 것이며^[5], 계층적 부호화 구조에 맞게 변형·적용되었음을 재차 언급한다.

비디오 영상 부호화의 RA 모드는 부호화 효율성이 높고 통신내러 상황에서도 복호화 영상 회복률이 높으며 무엇보다 사용자가 원하는 임의의 화면으로 손쉽게 접근할 수 있는 장점이 있어서 시장에서의 활용도가 크다. 즉 특정 프레임 임으로의 접근성이 좋고 대역폭 효율성이 높아 비디오 영상 편집이나 영상 스트리밍과 같이 대역폭 효율성이 요구되는 애플리케이션에 활용될 수 있다. 그림 1은 임의접근모드의 예측구조를 나타낸다. 그림에서 보는 바와 같이 RA 모드는 4개의 계층 구조를 가지고 있어 계층간의 정보를 활용할 수 있는 구조이다. I는 Intra-coded frame (I-프레임)을 의미하며 I-프레임은 화면내 정보를 이용하여 부호화된다. 아울러 B는 Bi-directionally coded frame (B-프레임)을 의미하며 과거 및 미래 프레임을 참조하여 부호화할 수 있

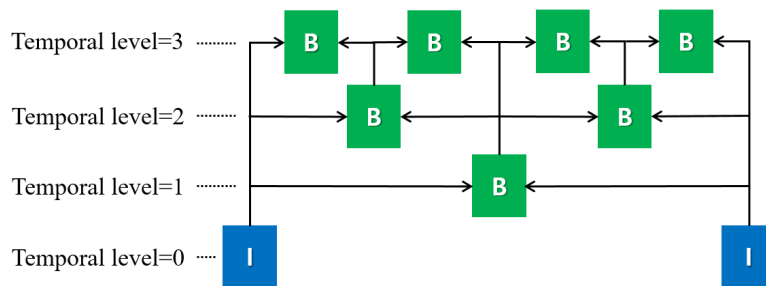


그림 1. 임의 접근 모드의 예측구조
 Fig. 1. A prediction structure of random access mode

는 특징이 있다. 뿐만 아니라 임의접근 모드에서는 계층적 B-프레임 구조를 사용하기 때문에 부호화 효율을 한층 더 높일 수 있는 기술적 특징이 있으며, 또한 주기적으로 I-프레임이 삽입되는 특성으로 인하여 특정 지점 복호화를 통한 임의접근이 가능하다는 큰 장점이 있다.

하지만 RA 모드의 경우 I-프레임이 전체 비디오 영상의 비트량 중 많은 부분을 차지하므로 비트율 정확도를 맞추기가 어렵고, 비트율-제어도 다른 예측모드에 비해 다소 어려운 단점이 있다. 이러한 단점들로 인해 발생하는 목표비트 비정확성은 통신 대역폭 자원의 낭비를 초래하고 부호화 효율성을 떨어뜨리게 된다. 따라서 RA 모드의 단점을 극복하여 목표비트율 정확도를 높일 수 있는 비트율-제어 알고리즘의 설계가 무엇보다 중요하다.

본 논문의 제안방법은 RA 모드에서의 목표비트율 정확도를 높이기 위해 RDO에 사용되는 λ 값을 효율적으로 예측하고 조정하는 방법으로서 이전 및 현재 GOP 안에 있는 동일 위치 프레임 간의 화질왜곡 차이 정보를 바탕으로 부호화 예정 프레임의 λ 값을 계산하여 부호화에 사용한다.

GOP 본 논문 제 II장 2절의 파티클 필터링 기반의 비트율-제어에서는 (5)의 화질왜곡 지표를 임의의 비트값 파티클을 예측·생성할 때의 보조 도구로 사용하였다. 반면 본 제안방법에서는 일정 프레임간의 부호화 된 비트값 차이의 비율을 임의의 화질왜곡 파티클 예측·생성의 보조 도구로 사용하였다. 해당 값은 다음의 수식으로 계산된다.

$$\Delta B_i = \beta \cdot (B_i - B_{i-j}) / B_{i-j} \quad (9)$$

여기서 B 는 실제 부호화 후에 얻어진 비트값을 의미하고, i 는 현재 프레임의 번호, $(i-j)$ 는 현재보다 j 번째 이전의 프레임 번호를 의미한다. β 는 비트율 조절 파라미터로 (5)의 수치와 동일한 수치를 사용하였다. (9)를 이용하여 생성된 다음 프레임의 화질왜곡 파티클은 다음과 같이 표현된다.

$$\bar{d}_{i+1} = D_i \cdot \underbrace{[1 \ 1 \ 1 \ \dots \ 1]}_N^T + \Delta B_i \cdot \bar{d}_i \quad (10)$$

여기서 \bar{d}_{i+1} 및 \bar{d}_i 는 벡터로 표현된 $(i+1)$ 및 i 번째 프레임의 N 개의 화질왜곡 파티클이며, D_i 는 비디오 영상 부

호화의 결과인 i 번째 프레임의 스칼라 평균제곱오차값, ΔB_i 는 (9)의 값이다. (10)의 화질왜곡 파티클들은 무작위 재추출 및 확률정규화 과정을 거쳐 다음의 수식으로 표현된다. 여기서 사용된 무작위 재추출 방법은 제 II장 2절의 방법과 동일한 방법이다.

$$q_{k,i+1} = p(d_{k,i+1}) / \sum_{l=1}^N p(d_{l,i+1}), \quad k = 1, 2, \dots, N \quad (11)$$

여기서 $p(d_{k,i+1})$ 는 \bar{d}_{i+1} 의 pdf이며, \bar{q}_{i+1} 는 확률정규화 과정을 거쳐 생성된 $(i+1)$ 번째 프레임의 화질왜곡 파티클(N 개)의 확률이다. 따라서 이렇게 생성된 (10)과 (11)의 화질왜곡 파티클 및 해당 확률값의 내적으로 최종 예측 화질왜곡값이 계산되며, 그 값은 다음의 수식과 같다.

$$\tilde{D}_{i+1} = \bar{q}_{i+1} \cdot \bar{d}_{i+1} \quad (12)$$

(12)로 예측된 최종 예측 화질왜곡값은 비디오 영상 부호화의 RDO에 사용되는 λ 를 세밀하게 조절하는 조정값 계산에 사용되며, 이는 다음의 수식으로 계산된다.

$$\gamma_{i+1} = \theta \cdot (\tilde{D}_{i+1} - D_{i-j}) / D_{i-j} \quad (13)$$

여기서 γ_{i+1} 는 $(i+1)$ 번째 프레임의 λ 값 조절을 위한 수치이며, j 는 RA 모드 GOP 구조의 I-프레임 주기와 같도록 8로 설정하였다. θ 는 화질왜곡 변화를 완만하게 조절하는 상수값으로서 실험적 수치로 0.03을 사용하였다. (13)을 비디오 영상 부호화의 RDO용 λ 조절에 사용하였고, 다음 프레임을 위한 부호화용 λ 는 다음과 같이 계산된다.

$$\lambda_{i+1} = (\gamma_{i+1} + 1) \cdot \lambda_i \quad (14)$$

비디오 영상 부호화 진행시의 급격한 비트율 변화를 방지하기 위하여 다음의 수식으로 비트율 변화의 정도를 원 수치 대비 $\pm 5\%$ 수준으로 제한하였고, 그 상하한의 범위는 다음과 같다.

$$\begin{aligned} \lambda_{i+1} &= 1.05\lambda_i ; \gamma_{i+1} \geq +0.05 \\ \lambda_{i+1} &= 0.95\lambda_i ; \gamma_{i+1} \leq -0.05 \end{aligned} \quad (15)$$

본 논문에서 고안한 (14)와 (15)의 λ 값 조절용 수식은 RA 모드에 적합한 예측구조를 갖도록 λ 값을 효율적으로 조절하였다. 이는 비디오 영상 부호화 비트율-제어 기법의 RDO 단계에 적용되어 기존 VTM-16.0rc1 기본 RC 대비 목표비트율 정확도 성능을 향상시켰다. 다음 제 IV장에서는 본 제안 알고리즘을 평가하기 위한 다양한 실험조건 및 결과를 정리·분석한다.

IV. 실험결과

1. 실험조건

제안방법의 공정한 성능평가를 위해, Joint Video Exploration Team (JVET)에서 권고하는 Common Test Conditions (CTC) 조건의 QP (=22, 27, 32, 37) 값으로 비디오 영상 부호화를 진행하였다. 실험에는 총 12개의 Class A (2560×1600), Class B (1920×1080), Class C (832×480), Class D (416×240), Class E (1280×720), Class A1 (3840×2160) 테스트 영상을 클래스 별로 2개씩 사용하였다. 또한 비디오 영상 부호화에 사용된 프레임은 각 테스트 영상 시퀀스의 초기 100 프레임을 사용하였다. 표 1은 실험에 사용된 테스트 영상 시퀀스의 특성을 보여준다.

비디오 영상 부호화 모드는 RA 모드를 사용하였으며 I-프레임 주기 및 GOP 값은 8을 사용하였다. 실험은 인텔 코어 i-7-8700K CPU (@3.7 GHz) 및 32.0GB RAM을 갖춘 64비트 윈도우 10 운영체제의 PC에서 수행되었다. 실험에

사용된 참조 소프트웨어 VVC Test Model (VTM)의 버전은 16.0rc1이며, 비디오 영상 부호화를 위한 블록 분할 구조에 사용된 Coding Tree Unit (CTU)의 크기는 128, Coding Unit (CU)의 가로 및 세로의 최대 크기는 64이다. 비디오 영상 부호화에 사용된 기본 부호화 기능으로서 Sample Adaptive Offset (SAO), TransformSkip, TransformSkipFast를 활성화하여 부호화하였고, 양자화 관련 부호화 기능으로서 Rate-Distortion Optimized Quantization (RDOQ), RDOQ for Transform Skip (RDOQTS)을 활성화하여 부호화하였다. 또한 비트율-제어 수행 수준을 picture-level 단계로 설정하였으며 부호화에 초기 QP값은 지정하지 않았다.

2. 정량평가

비트율-제어의 성능평가를 위해 다양한 평가지표가 사용될 수 있다. 본 논문에서 적용한 비디오 영상 부호화의 RA 모드는 비디오 영상 부호화를 위한 예측 구조상 계층적 예측구조의 형태를 갖기 때문에 주기적으로 I-프레임이 삽입되는 구조가 된다. 즉 이러한 구조 때문에 비트율-제어의 가장 기본적인 성능인 비트율 정확도를 맞추기가 어렵다. 따라서 본 논문에서는 통신 대역폭 자원의 낭비를 최소화하고 또한 초당 일정한 비트율을 생성하여 실시간 데이터 처리 효율성을 제고하는 것에 초점을 두었다. 따라서 성능평가의 지표로서 목표비트율 정확도 (bit-rate accuracy, BRA)를 사용하여 비트율-제어 성능을 평가하였다. 다음은 BRA 평가지표 계산 수식이다.

표 1. 실험에 사용된 테스트 시퀀스
 Table 1. Test sequences for experiments

Sequence	Resolution(W×H)	Bit-depth	Class	Frame per sec.	No. of frames
<i>BlowingBubbles</i>	(416×240)	8	D	50	100
<i>RaceHorses</i>	(416×240)	8	D	30	100
<i>BQMall</i>	(832×480)	8	C	60	100
<i>BasketballDrill</i>	(832×480)	8	C	50	100
<i>Kristen and Sara</i>	(1280×720)	8	E	60	100
<i>FourPeople</i>	(1280×720)	8	E	60	100
<i>Cactus</i>	(1920×1080)	8	B	50	100
<i>ParkScene</i>	(1920×1080)	8	B	24	100
<i>Traffic</i>	(2560×1600)	8	A	30	100
<i>People on Street</i>	(2560×1600)	8	A	30	100
<i>Tango2</i>	(3840×2160)	10	A1	60	100
<i>FoodMarket4</i>	(3840×2160)	10	A1	60	100

$$BRA(\%) = \left[1 - \frac{|BR_{target} - BR_{coded}|}{BR_{target}} \right] \times 100 \quad (16)$$

여기서 BR_{target} 은 목표비트율, BR_{coded} 는 실부호화 비트율을 의미한다.

표 2는 목표비트율 정확도 실험결과를 나타낸다. CTC 조건의 QP (=22, 27, 32, 37)을 가지고 비디오 영상 부호화

를 하되, 비트율-제어에 사용되는 목표비트율 계산을 위해 참조 소프트웨어 VTM의 RC 모드를 비활성화하여 부호화하였다. 제안방법의 성능평가를 위해 VTM-16.0rc1의 기본 RC 모드에서의 BRA와 제안방법의 RC 모드의 BRA를 비교하였다. 표 2의 결과에서 볼 수 있듯이 대부분의 QP에서 제안방법의 RC가 VTM-16.0rc1의 기본 RC 보다 우수한 성능을 나타내었다(BRA의 값을 비교하여 더 우수한 성능의

표 2. 목표비트율 정확도 실험결과

Table 2. Experimental results for bit-rate accuracy (BRA)

Sequence	QP	VTM-16.0rc1 RC off	VTM-16.0rc1 default-RC		Proposed RC	
		Target Bit-rates(Kbps)	Encoded Bit-rates(Kbps)	BRA (%)	Encoded Bit-rates(Kbps)	BRA (%)
BlowingBubbles	22	2745.252	1278.280	46.56	1399.360	50.97
	27	1393.204	817.904	58.71	966.748	69.39
	32	740.740	533.864	72.07	643.588	86.88
	37	381.340	337.680	88.55	403.752	94.12
RaceHorses	22	1710.540	798.46	46.68	854.246	49.94
	27	918.345	552.468	60.16	639.710	69.66
	32	474.124	348.408	73.48	449.973	94.91
	37	235.075	203.815	86.70	294.746	74.62
BQMall	22	8605.435	3897.868	45.30	4279.488	49.73
	27	4209.403	2536.881	60.27	2910.504	69.14
	32	2321.136	1746.590	75.25	2035.857	87.71
	37	1278.604	1160.932	90.80	1380.048	92.07
BasketballDrill	22	5833.904	2806.232	48.10	3142.088	53.86
	27	2830.792	1768.568	62.48	2056.004	72.63
	32	1432.384	1102.800	76.99	1334.784	93.19
	37	739.360	661.304	89.44	844.752	85.75
Kristen and Sara	22	5510.500	2748.355	49.87	3397.190	61.65
	27	2565.124	1706.515	66.53	2084.467	81.26
	32	1470.412	1176.681	80.02	1313.054	89.30
	37	860.947	796.992	92.57	828.993	96.29
FourPeople	22	6958.027	3526.910	50.69	4206.638	60.46
	27	3421.377	2267.222	66.27	2695.022	78.77
	32	2032.660	1616.654	79.53	1767.561	86.96
	37	1230.950	1143.163	92.87	1170.086	95.06
Cactus	22	36322.224	14588.7	40.16	16874.236	46.46
	27	11812.040	7353.692	62.26	9066.512	76.76
	32	5624.224	4399.200	78.22	5535.976	98.43
	37	2887.548	2650.264	91.78	3264.296	86.95
ParkScene	22	14203.441	6525.891	45.95	7608.017	53.56
	27	6470.327	3908.165	60.40	4939.167	76.34
	32	3152.432	2388.608	75.77	2980.416	94.54
	37	1509.394	1385.028	91.76	1709.195	86.76
Traffic	22	27691.238	13263.66	47.90	15672.439	56.60
	27	12503.592	8056.802	64.44	9953.889	79.61
	32	6480.880	5251.336	81.03	6157.528	95.01
	37	3492.067	2692.862	77.11	3585.120	97.34
People on Street	22	44220.645	20034.03	45.30	22225.670	50.26
	27	20893.972	13183.31	63.10	15247.046	72.97
	32	10972.324	8671.226	79.03	10845.590	98.84
	37	5819.457	5276.762	90.67	7560.000	70.09
Tango2	22	72564.604	33269.966	45.85	47645.971	65.66
	27	15837.249	13202.726	83.37	16613.246	95.10
	32	6825.057	6750.292	98.90	6750.292	98.90
	37	3722.577	3543.340	95.19	3543.340	95.19
FoodMarket4	22	18455.788	15075.465	81.68	19240.036	95.75
	27	7075.012	7437.849	94.87	7437.849	94.87
	32	3864.614	3979.425	97.03	3979.425	97.03
	37	2200.569	2100.868	95.47	2100.868	95.47

수치에 굵은 글꼴을 사용하여 표현하였다). 전반적으로 작은 QP를 사용하여 고화질로 영상을 부호화할 시에 목표비트율 정확도가 떨어지는 경향이 있었다. 이는 고화질 부호화시에 I-프레임에 많은 비트가 할당되기 때문으로 사료된다. 그러나 제안방법에서의 λ 값 조정은 부호화 비트할당을 효과적으로 수행하여 목표비트율 정확도를 높이고 있음을 실험결과로부터 확인할 수 있었다. 기존방법의 경우 Class A1의 4K 영상에서 대체적으로 목표비트율 정확도가 높은 특징이 있었다(정확도 90% 이상 달성). 따라서 해당 클래스 영상의 경우 목표비트율 정확도가 90% 미만일 경우에만하여 제안방법을 적용하여 결과를 도출하였다.

표 3은 실험에 사용된 각 테스트 영상 시퀀스의 BRA 평균값을 나타낸다. 이는 각 테스트 영상 시퀀스마다 4개의

QP로 비디오 영상을 부호화하여 얻은 BRA 평균값이다. 또한 각 테스트 영상 시퀀스의 전체 BRA 평균값은 VTM-16.0rc1의 기본 RC 모드의 경우 71.82%, 제안방법 RC의 경우 79.64%이다. 분석결과 제안방법은 VTM-16.0rc1의 기본 RC 모드보다 7.82% 포인트 BRA 성능이 개선되었음을 알 수 있다. 이는 제안방법이 RA 모드에 적합한 예측구조를 갖도록 λ 값을 효율적으로 조정하였고 그 λ 값이 RDO 과정에 효과적으로 적용되었기 때문이라고 판단된다. 결론적으로 제안방법 RC는 부호화 필요 비트를 더 정확히 생성하여 목표비트율 정확성을 향상시켰다.

그림 2는 제안하는 방법의 RC와 VTM-16.0rc1 기본 RC의 R-D 성능을 나타낸다. R-D 성능 그래프의 PSNR (peak signal to noise ratio)값은 Y값을 사용하였다. 제안하는 방

표 3. 평균 목표비트율 정확도

Table 3. Experimental results for an average bit-rate accuracy (BRA)

Sequence	VTM-16.0rc1 default-RC	Proposed RC
	Average Bit Rate Accuracy (%) with 4-QPs	
<i>BlowingBubbles</i>	66.47	75.34
<i>RaceHorses</i>	66.76	72.28
<i>BQMall</i>	67.90	74.66
<i>BasketballDrill</i>	69.25	76.36
<i>Kristen and Sara</i>	72.25	82.12
<i>FourPeople</i>	72.34	80.31
<i>Cactus</i>	68.11	77.15
<i>ParkScene</i>	68.47	77.80
<i>Traffic</i>	67.62	82.14
<i>People on Street</i>	69.53	73.04
<i>Tango2</i>	80.83	88.71
<i>FoodMarket4</i>	92.26	95.78
<i>Total Average</i>	71.82	79.64

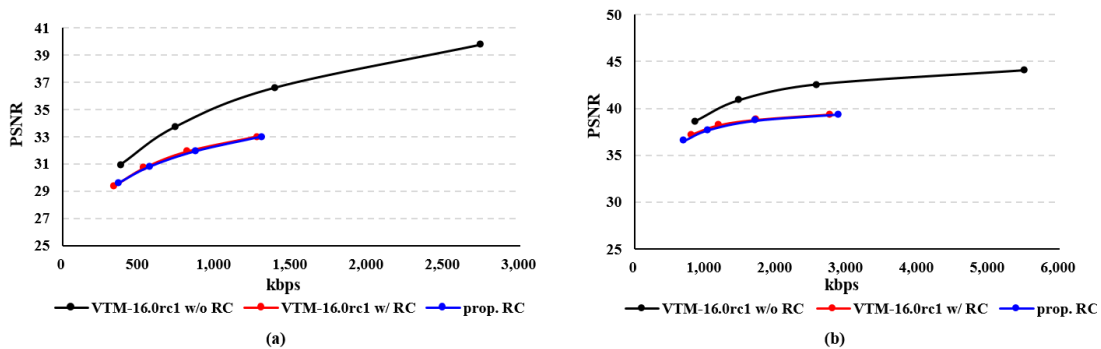


그림 2. R-D성능 비교 (a) BlowingBubbles (b) Kristen and Sara

Fig. 2. R-D performance comparisons (a) BlowingBubbles (b) Kristen and Sara

법의 RC와 VTM-16.0rc1 기본 RC를 비교했을 때 두 방법 간 R-D 성능에는 큰 차이가 없었다. 따라서 제안방법의 RC는 유의미한 R-D 성능의 변화 및 열화가 없으면서도 BRA를 높일 수 있는 방법이다.

3. 토의 및 고찰

비디오 영상 부호화 모드 중 장점이 많고 활용도가 높은 모드가 RA 모드인데, 그 이유는 비디오 영상 부호화의 GOP 중간중간에 I-프레임을 삽입하여 다양한 기능을 수행할 수 있기 때문이다. 하지만 I-프레임이 전체 비디오 영상의 비트량 중 많은 부분을 차지하므로 비트율 정확도를 맞추기가 어렵고 목표비트율 비정확성은 통신 대역폭 자원의 낭비를 초래하며 부호화 효율성을 떨어뜨리게 된다. 따라서 본 논문의 제안방법은 표 2와 표 3의 실험결과에서 보여주듯이, RA 모드의 단점인 목표비트율 정확도 성능을 크게 높였다는 점에서 기술적으로 큰 기여를 하였다. 하지만 다음의 몇 가지 부분은 후속 연구의 영역으로 추가 연구가 필요할 것으로 사료된다.

첫째로 표 2에서 볼 수 있듯이 QP를 작게 사용하여 고화질로 부호화하는 경우는 대체적으로 BRA 결과가 상대적으로 낮은 것을 알 수 있다. 즉, QP가 작아지면 BRA 성능도 떨어지는 것을 알 수 있다. 본 논문의 실험에서는 화질왜곡 변화를 완만하게 조절하는 상수값으로 (13)의 θ 를 사용하였다. 추후 QP와 θ 의 관계를 실험적으로 모델링하여 상관성을 찾아내 알고리즘에 적용한다면 작은 QP에서도 보다 더 향상된 BRA 성능을 기대할 수 있을 것이라 판단된다.

둘째로 (9)에서 (12)까지의 화질왜곡값을 이용한 파티클 필터링 과정은 제 II장 2절 (5)에서 (8)까지의 비트값을 이용한 파티클 필터링 과정과 유사하다. 즉 파티클을 어떠한 값으로 예측·생성할 것인가만 다를 뿐이다. 그러므로 파티클을 예측하는 정도를 화질왜곡 파티클에서는 비트변화율을, 비트값 파티클에서는 화질왜곡변화율을 사용함으로써 상호 예측 성능 향상에 보완 작용을 하고 있음을 알 수 있다. 이렇듯 예측정확성을 높이기 위해서는 서로 다른 속성의 파티클을 이용하여 예측을 각각 수행하고 유의미한 관계성 모델을 이용하여 서로 상보하는 방법이 효과적일 수 있다. 따라서 서로 다른 속성의 파티클을 이용한 예측 및

상호 연관성을 기반으로 한 예측 보정을 통해 예측성능을 향상하는 기법의 연구가 필요할 것으로 판단된다.

셋째로 본 연구에서는 RA 모드의 가장 기본적인 GOP 구조를 사용하여 실험을 수행하였다. 추후 조금 더 다양한 GOP 구조에서의 실험 수행을 통하여 각 구조별 장단점 분석을 한다면 제안 알고리즘의 완성도가 높아질 것으로 예상된다.

마지막으로 실험결과 분석에 있어서 본 제안방법의 RC에서는 MSE를 화질왜곡 정도의 지표로 사용하여 RDO과정을 수행하였다. 그리고 R-D 성능평가로 PSNR을 이용하였다. 본 논문에서는 시인성을 고려한 주관적인 화질 개선을 성능의 지표로 삼지 않았기 때문에 SSIM (structural similarity index measure)과 같은 지표는 사용하지 않았다. 아울러 본 논문의 참조소프트웨어로 VTM-16.0rc1 버전을 사용하였다. 이는 VVC 표준화 완료 버전 VTM 12.0 보다 최신 버전이다. 따라서 VTM-16.0rc1은 표준화에 적용된 기술들이 모두 적용된 참조소프트웨어로서 공정한 성능평가가 가능하였다고 판단된다.

V. 결론

본 논문에서는 비트율-왜곡 최적화 (rate-distortion optimization, RDO) 부분에 사용되는 λ 값을 조정하여 비트를 더욱 효과적으로 할당하는 방법을 제안하였다. λ 값 조정을 위하여 파티클 필터링 기반의 예측방법을 이용하였고, 이를 비디오 영상 부호화의 임의접근 (random access, RA) 모드에 적용하여 실험을 하였다. 실험결과 해당 방법이 계층적 구조의 RA 모드에서도 효과적으로 λ 값을 조정하고 있음을 확인하였다. 특히 목표비트율 정확도 성능평가 실험에서 제안방법이 기존방법 보다 평균 7.82% 포인트의 성능개선 효과를 달성하였음을 확인하였다. 따라서 본 연구의 알고리즘을 비트율-제어가 필요한 다양한 분야에 적용하면 전송 대역의 효율적 사용에 큰 이득이 있을 것으로 판단된다. 아울러 추후연구로서 다양한 GOP 구조별 성능장단점 분석을 통하여 제안 알고리즘의 완성도를 높일 예정이다.

참 고 문 헌 (References)

- [1] B. Bross, J. Chen, S. Liu, and Y. Wang, *Versatile video coding editorial refinements on draft 10*, 20th Meeting (Tele-conference), Oct. 2020.
- [2] B. Bross et al., "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736 - 3764, Oct. 2021.
doi: <https://doi.org/10.1109/TCSVT.2021.3101953>
- [3] M. H. Hyun, B. Lee, and M. Kim, "A novel rate and distortion estimation method using particle filtering based prediction for intra-predictive coding of deep block partitioning structures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 782 - 797, Feb. 2021.
doi: <https://doi.org/10.1109/TCSVT.2020.2989185>
- [4] M. H. Hyun, B. Lee, and M. Kim, "A frame-level constant bit-rate control using recursive bayesian estimation for versatile video coding," *IEEE Access*, vol. 8, pp. 227255 - 227269, 2020.
doi: <https://doi.org/10.1109/ACCESS.2020.3046043>
- [5] M. H. Hyun, B. Lee, and M. Kim, "A VVC intra rate control with small bit fluctuations using a Lagrange multiplier adjustment," *IEEE Trans. Multimedia*, vol. 26, pp. 6811-6821, Jan. 2024.
doi: <https://doi.org/10.1109/TMM.2024.3355633>
- [6] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston, MA, USA: Artech House, 2004.
- [7] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace distribution based lagrangian rate distortion optimization for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 193 - 205, Feb. 2009.
doi: <https://doi.org/10.1109/TCSVT.2008.2009255>
- [8] B. Li, H. Li, L. Li, and J. Zhang, " λ domain rate control algorithm for high efficiency video coding," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3841 - 3854, Sep. 2014.
doi: <https://doi.org/10.1109/TIP.2014.2336550>

저 자 소 개



현 명 환

- 2000년 3월 ~ 2005년 8월 : 인하대학교 정보통신공학부 학사
- 2004년 2월 ~ 2004년 7월 : City University of New York at Queens College, Flushing, New York 국가장학교환학생
- 2006년 3월 ~ 2008년 2월 : 광주과학기술원 정보기전공학부 석사
- 2008년 2월 ~ 2024년 2월 : 국방과학연구소(ADD) 미사일연구원 연구원/선임연구원/책임연구원
- 2016년 3월 ~ 2021년 2월 : 한국과학기술원 전기및전자공학부 박사
- 2024년 3월 ~ 현재 : 경상국립대학교 IT공과대학 제어로봇공학과 조교수
- ORCID : <https://orcid.org/0000-0002-8364-6127>
- 주관심분야 : 비디오 부호화, 비디오 비트율-제어, 영상처리, 컴퓨터비전, 인공지능기반 국방 및 항공우주 영상 시스템