

몰입형 콘텐츠를 위한 비트레이트 래더 생성

한건희¹⁾ 최경호¹⁾ 이승미¹⁾ 김순철²⁾ 오혜주²⁾ 정종범¹⁾

1) 국립군산대학교 컴퓨터정보공학과

2) 한국전자통신연구원

{gunhee9369, cgh3578, qweasd030622, jongbeomjeong}@kunsan.ac.kr

{choulsim, feeler}@etri.re.kr

Towards Bitrate Ladder for Immersive Contents

Geon Hee Han¹⁾ Gyeong Ho Choi¹⁾ Seung-Mi Lee¹⁾ Soonchoul Kim²⁾ Hyeju Oh²⁾ Jong-Beom Jeong¹⁾1) Department of Computer Science and Information Engineering,
Kunsan National University (KSNU)

2) Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문은 기존의 적응형 비트레이트 (adaptive bitrate streaming, ABR) 방식의 네트워크 혼잡 시 화질 저하를 통한 대역폭 문제 해결 기법 중 다중 해상도 부호화 파일 저장으로 인한 서버 저장공간의 한계를 극복하기 위한 방법을 제안한다. 콘텐츠의 유형에 따라 비트레이트 대비 video multi-method assessment fusion (VMAF) 점수기반의 비트레이트 래더 (bitrate ladder) 를 생성하여 각 영상 특성에 최적화된 양자화 매개변수 (quantization parameter, QP) 조합을 확률적으로 탐색한다. 결과적으로 기존 전수 조사 대비 높은 압축 효율을 달성하면서 사용자 체감 품질 (quality of experience, QoE) 을 유지할 수 있음을 확인하여, 사용자 관심영역 중심의 스트리밍 서비스 구현에 기여할 것으로 기대된다.

1. 서론

최근 비디오 스트리밍 서비스는 단순한 콘텐츠 소비를 넘어 전 세계 네트워크 구조를 결정짓는 핵심 요소로 자리 잡았다. 2025년 발표된 인터넷 현상 보고서에 따르면, Alphabet과 Meta 등 빅테크 기업들이 생성하는 트래픽은 모바일 네트워크의 35% 이상을 점유하며 인터넷 생태계에서 압도적인 비중을 차지하고 있다. 특히 Netflix와 Peacock 등 주요 OTT 플랫폼의 라이브 스포츠 독점 중계에서는 트래픽이 평소 대비 최대 60%까지 폭증하는 것으로 관측된다. 이러한 급격한 대역폭 변동은 네트워크 혼잡, 서비스 중단, 버퍼링 증가로 이어져 피크 트래픽 상황에 대한 효과적인 대안 마련이 필요하다 [1]. 기존 피크 트래픽 대응 방식은 적응형 비트레이트 기술에 주로 의존한다. 적응형 비트레이트 방식은 사용자와 서버의 가용 대역폭에 따라

영상 화질을 동적으로 조절하며, 이를 위해 서버에서는 원본 영상을 다양한 해상도로 다운스케일 (down-scale) 하여 부호화한 여러 버전의 파일을 저장한다. 클라이언트는 네트워크 상태를 실시간으로 확인한 후 최적의 해상도를 선택하여 영상을 수신한다. 그러나 해상도별로 다중 부호화 파일을 저장해야 하므로 막대한 서버 저장 공간과 부호화 연산량이 요구된다는 한계가 있다.

본 논문은 이러한 문제를 해결하기 위해 객관적 화질 평가 지표로 VMAF를 기반으로 rate-distortion curve (RD-curve)를 생성하고, just noticeable difference (JND) 구간을 활용한 확률적 탐색 기법을 적용하여 최적의 비트레이트 래더를 도출하고자 한다. 이를 통해 각 영상 콘텐츠의 특성에 최적화된 비트레이트 래더를 생성함으로써 기존 방식 대비 부호화 복잡도와 저장 공간을 크게 절감하면서도 사용자 체감 품질을 유지할 수 있는

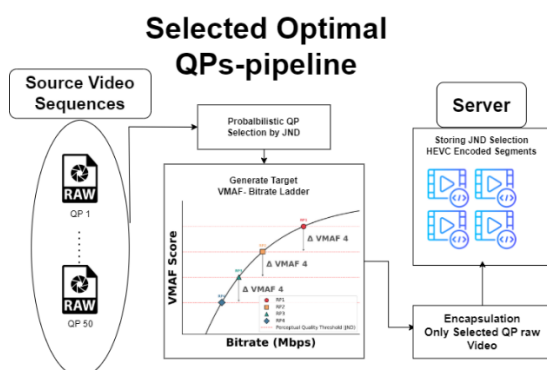
방안을 제시한다. 나아가 네트워크 혼잡 상황에서도 사용자 관심 영역(region of interest, RoI) 중심의 효율적인 스트리밍 서비스를 구현하기 위한 기술적 토대를 제공하고자 한다.

본 논문의 구성은 다음과 같다. 2절에서는 기존 적응형 비트레이트 기술 및 다중 해상도 캡슐화 과정에 관한 선행 연구를 소개한다. 3절에서는 VMAF 지표 기반의 JND를 추출하여 확률적 최적 비트레이트를 탐색 기법을 제안하고, 실험을 통해 자원 절감 효율성을 검증 및 분석한다. 마지막으로 4절에서는 본 논문의 결론을 맺는다.

2. 선행 연구

현재 비디오 스트리밍 서비스의 주류를 이루는 HTTP live streaming (HLS) 및 dynamic adaptive streaming over HTTP (DASH) 는 가용 네트워크 대역폭에 따라 화질을 동적으로 조절하는 적응형 비트레이트 기술을 핵심으로 한다. 이를 위해 서버는 하나의 원본 영상을 다양한 해상도 및 비트레이트 조합으로 부호화한 세그먼트들의 집합인 비트레이트 래더를 사전에 생성하여 저장한다. Apple HLS 권장 부호화 프로필에 따르면, 상용 서비스는 416×234부터 1920×1080까지 총 9개의 서로 다른 비트레이트 단계를 정의하며, 이 중 7개의 해상도를 주로 활용한다. 그러나 다중 해상도 구성 시 원본 영상 대비 약 2배 이상의 저장 공간을 점유하게 된다 [2]. 이러한 대용량 저장 공간 요구사항은 대규모 라이브 스트리밍 환경에서 서버 운영 비용을 크게 증가시키는 주요 요인으로 작용한다. 또한, 기존 방식은 영상 콘텐츠의 복잡도를 고려하지 않고 획일적인 비트레이트 래더를 적용한다. 이로 인해 배경이 단순한 뉴스나 토크쇼에서는 불필요하게 높은 저장 비용이 발생하고, 복잡한 움직임이 많은 액션 장면이나 스포츠 콘텐츠에서는 충분한 화질을 제공하지 못해 사용자 체감 품질 불균형이 발생한다. 따라서 기존의 다중 해상도 저장 방식은 서버 자원 고갈의 병목 현상을 초래하며, 콘텐츠 특성과 사용자 시각 인지 특성을 고려한 새로운 비트레이트 래더 최적화 기법이 요구된다 [3].

3. 본론



<그림 1> 선택적 부호화 파이프라인

본 논문은 기존 적응형 비트레이트 서비스의 다중 해상도 저장 방식을 개선하기 위해, 사용자의 시각 인지 임계치를 고려한 그림 1의 선택적 부호화 파이프라인을 제시한다. 해당 파이프라인은 기존 방식과 동일하게 입력으로 각 양자화 매개변수 QP에 따라 부호화된 파일들을 입력 받는다. 제안 방식은 VMAF 점수를 활용한 체감 화질 비트레이트 래더를 생성하여 JND를 측정한다. 기존 방식이 가용 대역폭에만 의존하여 모든 해상도의 세그먼트를 사전 생성하는 것과 달리, 제안하는 방식은 부호화 단계에서 확률적 QP 탐색 (probabilistic QP selection) 과정을 거친다. 기존 PSNR, structural similarity index (SSIM)은 사용자 체감 화질과의 상관관계가 낮다는 한계가 있다. 이를 극복하기 위해 Netflix에서 제안한 다중 화질 측정 모델 결합 지표인 VMAF를 선정하였다 [4, 5]. VMAF는 시각적으로 정보 손실 모델 (visual information fidelity, VIF), 디테일 손실 지표 (detail loss metric, DLM) 프레임 간의 움직임 변화량을 통합하여 점수를 산출한다. 본 논문은 VMAF 점수를 사용자의 체감 화질을 기준으로 JND의 단계를 나누기 위해 VMAF의 임계값을 총 3 구간 (2, 4, 6)으로 설정하였다. 본 논문의 제안 기법을 검증하기 위해 *Ultra Video Group* 데이터셋에서 *CityAlley*, *RaceNight*, *FlowerChild*, *HoneyBee*를 데이터셋으로 선정하였다.[6]



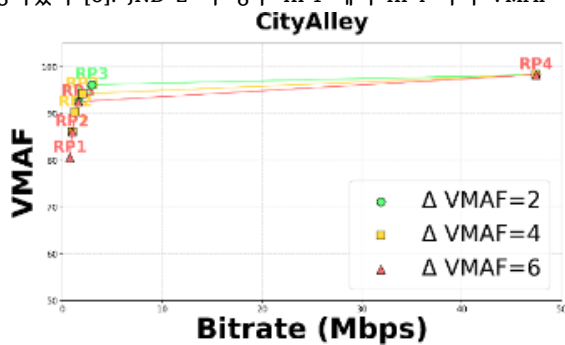
<그림 2> CityAlley 시퀀스의 JND 단계에 따른 특정 장면 화질 비교

각 시퀀스는 서로 다른 복잡도의 움직임을 가지며, HoneyBee 시퀀스만 120FPS, 나머지 시퀀스들은 50FPS로 동일 주사율이다. high-efficiency video coding (HEVC/H.265) 고속 구현체인 Kvazaar를 사용하여 QP 1~50 범위로 부호화를 수행하였고, VMAF (version 3.1.0) 모듈을 사용하여 VMAF를 측정하였다 [5, 7]. 그림 2는 CityAlley 시퀀스의 특정 장면을 JND 임계값별로 RP 단계에 따라 화질 변화를 비교한 결과이다. RP1에서 RP2로 이동할 때, JND 2에서는 오른쪽 인물의 선글라스 착용 여부를 확인할 수 있는 수준의 화질을 유지하였으나, JND 4와 JND 6에서는 얼굴 부분의 화질이 급격히 저하되었다. RP2에서 RP3으로 진행할 경우 JND 2에서도 얼굴 형태 구분이 어려워지는 화질 저하가 관찰되었으며, JND 6에서는 더욱 심각한 왜곡이 나타났다.

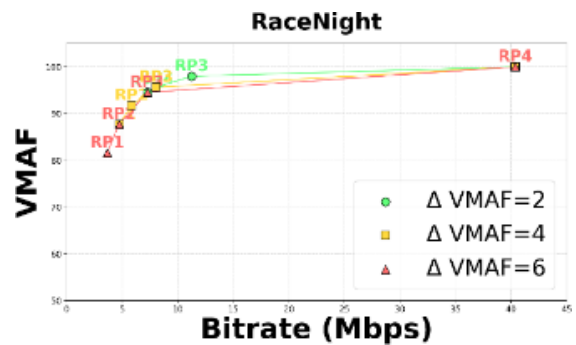
<표 1> CityAlley 시퀀스 VMAF SCORE 측정치

	RP 1	RP 2	RP 3	RP 4
JND 2	98.31	96.59	94.88	92.52
JND 4	98.31	94.88	90.33	87.64
JND 6	98.31	92.52	87.64	80.53

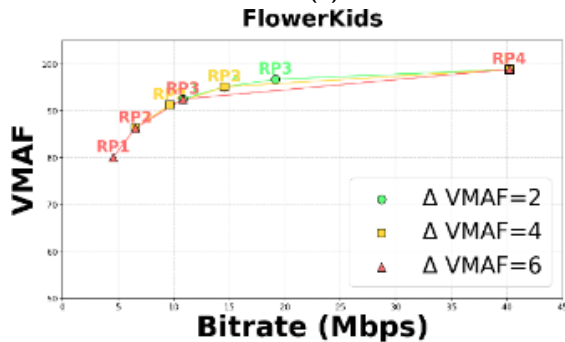
표 1은 위 그림 2의 각 단계별 이미지에 대한 VMAF 측정치를 정리한 것이다. VMAF 점수 최적 부호화 구간은 93 점으로 설정하였으며, 이하에서는 사용자 체감 화질 저하가 발생한다고 가정하였다 [8]. JND 2의 경우 RP1에서 RP4까지 VMAF 점수가



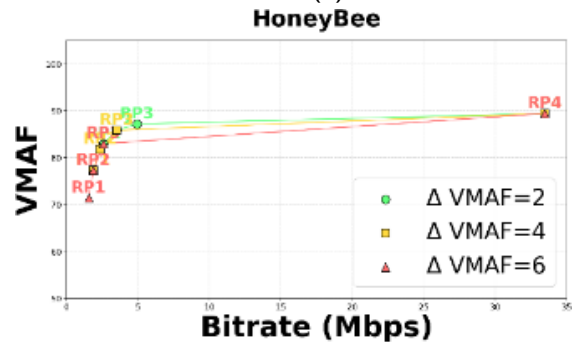
(a)



(b)



(c)



(d)

<그림 3> 데이터 시퀀스별 Bitrate 대비 VMAF RD-Curved 그래프

그림 3은 CityAlley 테스트 시퀀스의 특정 화면을 각 JND 임계점(2, 4, 6)별로 RP 1~4를 기준으로 화질 저하가 발생하는 비교군을 나열하였다. 그림 3은 CityAlley 시퀀스의 특정 화면을 JND 2, 4, 6별 RP 1~4로 비교한 결과이다. RP1에서 RP2로 이동 시 JND 2는 선글라스 착용 여부를 확인할 수 있는 수준을 유지하나, JND 4와 6에서는 얼굴 화질이 급격히 저하되었다. RP2에서 RP3으로 진행할 경우 JND 2와 JND 6에서도 얼굴 형태 구분이 어려워지는 화질 저하도 관찰되었다.

98.31에서 92.52로 완만하게 감소하여 체감 가능한 화질 저하가 거의 발생하지 않았다. 반면 JND 4와 JND 6에서는 RP3 이후 93점 이하로 떨어지며 체감 화질 저하가 시작되는 것으로 나타났다. 특히 JND 6은 VMAF 점수 편차가 가장 크며, 화질 저하가 급격하게 진행됨을 확인할 수 있었다. 이 결과는 JND 임계값이 클수록 압축 효율은 높아지지만, VMAF 점수 감소폭도 커진다는 트레이드오프 (trade-off) 관계를 보여준다. 따라서 적절한 JND 임계값 선정이 화질과 압축 효율의 균형을 결정하는 중요한 요소임을 알 수 있다.

<표 2> JND 별 점수 비용 대비 압축 효율

	감소된 VMAF 점수	절감된 Bitrate (Mbps)	VMAF 압축효율(η_k)
JND 2	5.79	1.37	0.236
JND 4	10.67	1.02	0.095
JND 6	17.78	0.84	0.047

표 2의 VMAF 1 점당 압축 효율은 부호화 과정에서 발생하는

시청자의 주관적 화질 저하 대비 대역폭 절감 효과의 상호 교환(trade-off) 관계를 정량적으로 평가한다. η_k 는 VMAF 1점당 압축효율을 뜻하고 R_{orig} 와 V_{orig} 는 각각 압축을 수행하지 않은 원본 비디오 시퀀스의 비트레이트 및 VMAF 점수를 나타낸다. R_k 와 V_k 는 k 번째 JND 단계의 부호화 R_k 와 V_k 는 k 번째 JND 단계의 부호화 제약 조건을 적용한 후 측정된 비트레이트, VMAF 점수를 의미한다. 변수 k 는 본 연구에서 설정한 JND 2, 4, 6 단계를 의미한다.

$$\eta_k = \Delta \frac{R_k}{V_k} = \frac{(R_{orig} - R_k)}{(V_{orig} - V_k)}, (1)$$

표 분석에 의하면 JND 2는 점수당 압축효율에서 높은 효율을 보여주므로 가장 압축 효율이 좋은 구간으로 측정된다. 또한 JND 4는 점수 1점당 대역폭을 1Mbps 이하로 감소시킬 수 있기 때문에 대역폭이 낮은 경우 좋은 선택지가 될 수 있다. 본 실험 결과를 종합하면, JND 임계치가 낮을수록 화질 보존력이 우수하며, 임계치가 높을수록 압축 효율이 극대화되는 전형적인 트레이드오프 관계를 확인하였다. 그러나 주목할 점은 희생된 화질 대비 비트레이트 압축효율이다. 이는 *CityAlley* 같이 인물의 이목구비나 질감이 중요한 시퀀스에서 시각적 품질을 저해하지 않고도 최적의 전송 효율을 낼 수 있는 임계점인 것을 보여준다 [9].

5. 결론

본 논문은 기존 적응형 비트레이트 기법의 서버의 저장공간을 낭비하는 한계를 극복하고 사용자 시각 인지 임계점인 JND를 사용하여 확률적 QP 탐색 기법을 사용하여 체감 화질을 기반으로 한 비트레이트 래더 생성 프레임워크를 제시한다. 본 연구는 JND를 2, 4, 6으로 설정하여 적응형 비트레이트 래더 생성시 확률적으로 탐색하여 시간 복잡도를 감소시킨다. PSNR같은 물리적 수치 대신 VMAF라는 사용자 체감 품질 수치를 적용하여 체감품질의 하락폭을 측정하였고 압축률과 VMAF 점수 간의 JND 상관관계를 평가하였다.

이러한 성과는 비디오 스트리밍 서비스의 저장공간 자원을 절약할 뿐만 아니라 서버의 연산량을 낮춰 처리속도와 서버의 부하량을 낮추는데 기여하였다. 또한, 추가 연구를 통해 전체 화면이 아닌 화면 내 특정 관심 영역만을 분할하여 처리하여 특정 타일만의 화질을 보존하고 나머지 타일의 화질을 저하하여 관심 영역의 품질을 유지하는 부분 압축 기법을 정립할 계획이다.

ACKNOWLEDGEMENT

이 논문은 2026년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구 결과임 (No. RS-2026-25511848, 클라우드기반 사용자참여형 오픈 미디어서비스 플랫폼 기술 개발)

참고 문헌

- [1] K. Rosenthal. The 2025 Global Internet Phenomena Report. Technical Report, Sandvine/AppLogic Networks. 2025.
- [2] A. Premkumar, P. T. Rajendran, V. V. Menon, A. Wieckowski, B. Bross, D. Marpe. Quality-Aware Dynamic Resolution Adaptation Framework for Adaptive Video Streaming. 15th ACM Multimedia Systems Conference, pp. 292-298. 2024.
- [3] Y. A. Reznik, K. O. Lillevold, A. Jagannath, J. Greer, J. Corley. Optimal design of encoding profiles for ABR streaming. 23rd Packet Video Workshop, pp. 43-47. 2018.
- [4] A. B. Watson. Toward a perceptual video-quality metric. Human Vision and Electronic Imaging III, SPIE, Vol. 3299, pp. 139-147. 1998.
- [5] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, M. Manohara. Toward a practical perceptual video quality metric. Netflix Tech Blog. 2016.
- [6] A. Mercat, M. Viitanen, J. Vanne. UVG dataset: 50/120fps 4K sequences for video codec analysis and development. *Proceedings of the 11th ACM Multimedia Systems Conference*, pp. 297-302. 2020.
- [7] A. Lemmetti, M. Viitanen, A. Mercat, J. Vanne. Kvazaar 2.0: fast and efficient open-source HEVC inter encoder. 11th ACM Multimedia Systems Conference, pp. 237-242. 2020.
- [8] R. Rassool, J. DeCock, A. Aaron. VMAF Reproducibility: Validating a Perceptual Practical Video Quality Metric. IEEE/SMPTE Motion Imaging Journal, Vol. 126, No. 7. 2017.
- [9] H. Wang, W. Gan, S. Hu, J. Han, S. K. Kattula, C.-C. J. Kuo. VideoSet: A large-scale compressed video quality dataset based on the JND concept. Journal of Visual Communication and Image Representation, Elsevier, Vol. 46, pp. 292-302. 2017.