

## 비디오 화질 자동 측정 기술 개발을 위한 데이터 셋 구축 방법

정세윤, 이대열, \*정연수, \*김태화, 조승현, 김휘용  
한국전자통신연구 실감 AV 연구그룹, \*㈜큐랩  
jsy@etri.re.kr

### Data Set Design Method for developing Automatic Video Quality Measurement Technology

Se Yoon Jeong, Dae Yeol Lee, \*Yeonsoo Jeong, \*Tae Hwa Kim, Seunghyun Cho, and Hui Yong Kim  
ETRI Realistic AV Research Group, \*Quram

#### 요 약

기계학습 기반 비디오 화질 자동 측정 기술은 주관적 화질 평가를 대체하기 위한 기술로, 비디오를 입력 신호로 화질 평가 결과를 출력 신호로 하는 기계학습 모델을 통해서 개발하는 기술이다. 학습에 필요한 비디오 데이터 셋은 입력 신호인 비디오 시퀀스와 입력의 출력신호로 학습할 주관적 화질 평가 결과로 구성된다. 이때 데이터 셋의 일부는 기계학습 기반 비디오 화질 자동 측정 기술 개발 과정에서 학습에 사용하고, 남은 일부는 개발 기술의 성능 평가에 사용한다. 일반적으로 기계학습 기반 기술의 성능은 학습 데이터의 양과 질에 비례한다. 그러나, 기계학습 기반 비디오 화질 자동 측정 기술 개발에 필요한 데이터 셋은 주관적 화질 평가 결과를 포함해야 하므로, 데이터 양을 늘리는 것은 쉬운 문제가 아니다. 이에 본 논문에서는 압축 비디오에 대한 화질 자동 측정 기술 개발을 위해 필요한 데이터 셋을 양과 질적 측면에서 효율적으로 구축하는 방법을 제안한다. 양적 측면에서 효율성을 높이기 위해 부호화 복잡도와 평가 난이도 기반으로 시퀀스를 선정 방법을, 질적 측면에서 효율성을 높이기 위해 쌍 비교(Pairwise Comparison)기반의 주관적 화질 평가 방법을 제안한다.

#### 1. 서론

비디오 부호화 응용에서 화질 척도로 가장 널리 사용되는 것은 PSNR 이다. 그러나, 객관적 화질 척도인 PSNR 은 인간이 인지적 화질과의 상관 정확도(correlation)가 높지 않기 때문에, 대부분의 응용서비스에는 인간이 직접 화질을 평가하는 주관적 화질 평가를 필요로 하고 있다. 주관적 화질 평가라는 명칭에서 알 수 있듯이 각 평가자의 화질에 대한 주관적 의견들을 측정하고 이를 통계적으로 평가하는 방법이다. 이때 통계적 결과 표현으로는 MOS(Mean Opinion Score)를 주로 사용한다. 통계적으로 신뢰성 있는 결과를 위해서 평가자의 수가 충분해야 한다. ITU-R. BT.500 [1]표준에서는 평가자의 수를 최소 15 명(권고로는 20 명 이상)으로 하고 있다. 또한 한번에 연속적으로 평가 하는 세션의 평가 시간은 30 분 이내로 하도록 규정하고 있고, 충분한 휴식 후 후속 평가 세션을 진행하도록 하도록 하고 있다. 이처럼 주관적 화질 평가는 많은 인력과 시간을 필요로 하므로, 주관적 화질 평가를 대체할 수 있는 자동 화질 측정 기술이 요구되고 있다.

자동 화질 측정 기술은 주관적 화질 평가를 대체할 수 있는 객관적 화질 척도를 개발을 포함하며 이를 위해서는 주관적 화질 평가 결과와 높은 상관도를 제공해야 한다. 자동 화질 측정 기술은 주관적 화질 평가 결과를 기계 학습을 통해 개발되고 있다. 본 논문에서 압축 비디오에 대한 화질을 자동 측정하는 기술에 대해서만 고려한다.

기계학습 기반의 압축 비디오 자동 화질 측정 기술을 개발

하는데 있어서 비디오 시퀀스와 대응 주관적 화질 평가 쌍으로 구성된 데이터 셋이 미리 구축되어야 한다. 일반적으로 기계학습에서는 학습 데이터의 양과 질에 비례하여 학습 결과가 향상되므로, 양과 질적 측면에서 좋은 데이터 셋을 구축하는 것이 중요하다. 데이터 셋 구축은 주관적 화질 평가를 동반하므로 주관적 화질 평가 시간을 줄이면서도 양적 질적 측면에서 효율적으로 데이터 셋을 구축하는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2 절에서는 기계학습 기반 비디오 화질 자동 측정 기술과 관련 있는 공개된 데이터 셋 동향을 통해 데이터 셋 구축에서 고려해야 할 사항들에 대해 설명한다. 3 절에서는 제안 데이터 셋 구축 방법에 대해 설명하고, 4 절에서는 구축한 사례를 소개하고, 5 절에서는 향후 연구 계획과 결론을 기술한다.

#### 2. 비디오 데이터 셋 동향

기계학습 기반의 압축 비디오 자동 화질 측정 기술에 활용할 수 있는 데이터 셋으로는 Texas Austin 대학의 LIVE (Laboratory for Image and Video Engineering) 데이터 셋 [2]과 VQEG (Video Quality Expert Group)의 데이터 셋 [3]이 있다. 이 2 기관의 데이터 셋은 BT.500[1]기반의 주관적 화질 평가 방법으로 측정된 MOS 값을 주관적 화질 평가 결과로 제공하고 있다.

USC 는 VideoSet [4]을 2017 년에 공개하였는데, 이 데이터 셋은 쌍 비교(Pairwise Comparison, 이하 PC) 기반의 동적

주관적 화질 평가를 통해서 측정된 JND 값들을 주관적 화질 평가 결과로 제공하고 있다. 이 데이터 셋은 220 개의 시퀀스를 4 개의 공간해상도(1080p, 720p, 540p, 360p)로 변환하고, 각 시퀀스를 AVC 로 QP 0~51 로 압축한 45,760 개의 압축 비디오 클립으로 구성되어 있다. 그리고, 화질 평가 결과는 각 압축 클립 별로 있는게 아니라 각 시퀀스 별로 첫번째~세번째 JND 위치 값을 제공하고 있다. 이때 JND 는 75% SUR[5] 을 사용하고 있다.

상기 동향으로 부터 데이터 셋을 포함한 주관적 화질 평가 정보에 따라 기존의 BT.500 기반의 MOS 계열과 최근에 제안된 PC 기반의 SUR 계열로 분류할 수 있다.

PC 기반의 주관적 화질 평가는 BT.500 기반의 주관적 화질 평가와 비교하여 보다 신뢰도 높은 화질 평가 결과를 얻을 수 있는 장점을 가지므로, 기계학습 측면에서 질적으로 보다 좋은 데이터 셋을 구축할 수 있다. 따라서 본 논문에서는 PC 기반의 주관적 화질 평가 결과를 기반으로 데이터 셋을 구축하는 방법을 제안한다.

### 3. 제안 데이터 셋 구축 방법

본 논문에서는 압축 비디오에 대한 화질을 자동 측정 기술을 개발하는데 양과 질적 측면에서 효율적인 데이터 셋을 구축하는 방법을 제안한다.

제안 데이터 셋 구축 방법을 VideoSet 과 유사하게 AVC QP 0~51 경우의 압축 비디오 클립과 각 시퀀스 별로 SUR 정보를 제공하는 것을 구축하는 경우를 통해 설명한다.

제안 데이터 셋 구축에서 첫 번째 고려한 사항은 필요한 평가 문항 수를 줄이는 것이다. PC 방법은 심리학에서 사용하는 방식으로 N 개의 항목이 있을 때 N(N-1) 개의 비교를 요구한다. 이에 비해 BT.500 의 대표적인 주관적 화질 평가인 DSCQS 의 경우 원본과의 비교 만을 수행하므로 N-1 개의 비교를 요구한다. 따라서, PC 기반의 주관적 화질 평가에서 필요한 평가 문항 수를 줄이는 것이 필요함을 알 수 있다. 이는 사전 지식을 활용하면 평가 문항 수를 크게 줄일 수 있다[6]. 본 데이터 구축에서 활용한 사전 지식은 QP 값이 증가할수록 화질은 떨어진다는 것이다. 이는 SUR 측정에 있어서, 평가자가 특정 QP 에서 화질 차이를 느꼈다면 특정 QP 보다 큰 QP 에서도 화질 차이를 느낄 가능성이 높다는 것이다. 이 정보를 활용하기 위해 VideoSet 의 이전 탐색과 유사한 QP 간격을 동적으로 조절하는 동적 화질 평가를 제안한다. 제안 방법의 핵심은 현재 기준 QP 에 대해 첫 번째로 화질 차이를 인지하는 QP 를 빨리 찾는 것이다. 이때 주관적 화질 평가에는 휴먼 에러가 존재할 수 있으므로 이를 고려하는 것도 중요하다. 따라서 첫 번째 화질 차이 QP 를 발견하면 바로 한단계 큰 QP 에 대해 비교를 수행해서 휴먼에러인지를 판별하고 판별이 유효하면 이후 큰 QP 들에 대해서는 평가를 수행하지 않도록 하여 비교 문항 수를 줄였다. 이를 화질 평가에 적용하기 위해서 그림 1 과 같은 화질 평가 프로그램을 구현하였다.



(a)평가 항목 재생 화면 (b)평가 결과 입력 화면  
 그림 1. 데이터 셋 구축을 위한 화질 평가 프로그램

제안 데이터 셋 구축에서 고려한 두 번째 사항은 적은 시퀀스로 높은 성능의 기계학습 효과를 높이기 위해 시퀀스를 분

류 통해 효율적으로 시퀀스를 선정하는 방법이다. 기계학습 효과를 높이기 위해서는 다양한 경우가 데이터 셋에 포함이 되어야 한다. 이를 위해 압축 비디오 화질 측정 기술 개발 측면을 고려하여 시퀀스를 부호화 복잡도와 화질 평가 난이도로 분류하는 방법을 제안한다.

부호화 복잡도는 비트율을 기준으로 분류를 하는 것이다. 일 예로 1080pHD 의 경우 QP 22 의 압축 비디오 클립의 비트율이 18Mbps 이상인 경우에는 상, 6~18Mbps 는 중, 6Mbps 이하는 하로 부호화 복잡도를 구분하였다.

화질 평가 난이도는 주관적 평가에서 평가자 간에 의견 차이가 클수록, 즉 분산의 크기가 클 수록 난이도가 높다고 생각하였다. 그러나, 주관적 평가 이전에 시퀀스를 분류하는 것이 필요하므로, 각 평가자 대신 여러 객관적 척도(SSIM, VQM-VFD, VMAF)로 측정된 화질 값들의 유사도를 차이 정도를 통해 분류하였다.

### 4. 데이터 셋 구축 사례

제안한 구축 방법을 통해 HD1080p 27 개 시퀀스에 대해 AVC QP 0~51 까지 압축 클립과 각 시퀀스에 대한 20 명의 SUR 화질 데이터를 포함하는 데이터 셋을 구축하였다. 제안 데이터 셋 구축 방법의 효율성을 판별하기 위한 분석은 현재 진행 중에 있다.

### 5. 결론

기계학습 기반 압축 비디오 화질 자동 측정 기술 개발을 위해서 데이터 셋 구축이 선행되어야 한다. 양과 질적 측면에서 효율적으로 데이터 셋을 구축하는 방법을 제안하였다. 제안한 방법은 기계 학습에 적합한 SUR 화질 평가 정보를 포함하고, 부호화 복잡도와 화질 평가 난이도 분류를 통해 효율적으로 시퀀스 선정하는 방법과 동적으로 PC 화질 평가를 수행하여 평가 시간을 줄이는 방법을 특징으로 한다. 향후 연구로는 제안한 데이터 셋의 효율성을 검증하기 위한 분석 연구와 데이터 셋의 추가 구축을 통해 기계학습 기반의 비디오 화질 자동 측정 기술을 개발하는 것이다.

### 감사의 글

이 논문은 2018 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발)

### 참고문헌

[1] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures", Jan. 2012.  
 [2] LIVE DataSet:  
<http://live.ece.utexas.edu/research/Quality/index.htm>  
 [3] VQEG DataSet:  
<https://www.its.bldrdoc.gov/vqeg/downloads.aspx>  
 [4] USC Videoset :  
<https://iee-dataport.org/documents/videoset>  
 [5] Q. Huang, H wang, S. Lim, H. Kim, S. Jeong, and C. Cuo, "Measre and Prediction of HEVC Perceptually Lossy/Lossless Boundary QP Values", DCC 2017, Apr. 2017.  
 [6] R. Song, Y. Li, Y. Jia, Y. Wang and P. Rao, "Efficient, robust and divisible paired comparison for subjective quality assessment", Multimedia Tools Application, Jul. 2016.