

초해상화 이미지 데이터셋 기반 최신 화질 평가 지표 분석

김용록, 신준하, 고현석*

*한양대학교

helloyr12@hanyang.ac.kr, ipip0114@hanyang.ac.kr, *hyunsuk@hanyang.ac.kr

Analysis of State-of-the-Art Quality Assessment Metrics on a Super-Resolution Image Dataset

Yongrok Kim, Junha Shin, Hyunsuk Ko*

*Hanyang Univ.

요약

고해상도 영상 콘텐츠의 수요 증가와 함께, 저해상도 영상을 고해상도로 변환하는 초해상화(Super-Resolution) 기술의 연구가 활발히 이루어지고 있다. 이러한 기술의 성능을 신뢰성 있게 평가하기 위해서는 적절한 영상 화질 평가(Image Quality Assessment, IQA) 지표의 역할이 필수적이다. 최근 다양한 최신 IQA 지표들이 초해상화 영상의 품질 특성을 보다 정교하게 반영하기 위해 제안되고 있다. 본 논문에서는 초해상화 영상 데이터셋을 기반으로 대표적인 최신 IQA 지표들의 성능을 체계적으로 비교·분석하고, 각 지표가 갖는 강점과 한계를 고찰한다. 이를 통해 향후 초해상화 영상 화질 평가 지표 개발에 필요한 방향성과 개선점을 제시한다.

I. 서론

고해상도 디스플레이와 방송·스트리밍 서비스의 확산으로 저해상도 방송 콘텐츠를 고해상도로 변환하기 위한 초해상화(Super-Resolution, SR) 기술의 활용이 증가하고 있다. 최근 딥러닝 기반 초해상화 기법들은 단순한 해상도 향상을 넘어 잡음 제거와 선명도 개선을 동시에 수행한다. 그러나 기존 영상 화질 평가(Image Quality Assessment, IQA) 기법들은 주로 왜곡의 정도를 측정하는 데 초점을 두고 있어, 초해상화 영상에서 나타나는 품질 향상 효과를 충분히 반영하지 못하는 한계가 있다. 본 논문에서는 저해상도 방송 콘텐츠에 초해상화 기법이 적용된 영상에 대한 화질 평가 분석을 위해 SREB 데이터셋[1]을 활용하여 최신 IQA 지표들의 성능을 비교·분석하고, 초해상화 영상 화질 평가에서의 한계를 분석하였다.

II. 데이터셋 및 실험 조건

IQA 기법의 성능 평가를 위해 SREB 데이터셋을 사용하였다. SREB 데이터셋은 1995년부터 2007년 사이의 SBS 방송 영상에서 추출된 30 장의 원본 저해상도 이미지로 구성되며, 블러와 잡음 등 실제 방송 환경에서 발생하는 왜곡을 포함한다. 각 이미지는 사진 다운샘플링 없이 7 종의 초해상화 기법과 두 가지 배율($\times 2$, $\times 4$)을 적용하여 총 420 장의 초해상화 영상이 생성되었으며, 각 초해상화 영상에 대해 주관적 화질 평가를 통해 획득된 화질 점수(Mean Opinion Score, MOS)가 함께 제공된다.

본 실험에서는 비참조(No-Reference IQA, NR-IQA) 및 부분참조(Reduced-Reference IQA, RR-IQA) 지표의 성능을 비교하였다. 비교를 위해 4 종의 NR-IQA 기법(BRISQUE [2], SRmetric [3], KLTSRQA [4], ARNIQA [5])과 2 종의 RR-IQA 기법(PSCT [6], PFIQA [7])을 사용하였으며, RR-IQA의 경우 원본 저해상도 영상을 참조로 활용하였다.

성능 평가는 예측 점수와 주관적 화질 점수(MOS) 간의 Pearson 선형 상관계수(PLCC), Spearman 순위 상관계수(SRCC), 그리고 Root Mean Square Error(RMSE)를 측정하였다. 전체 데이터셋은 원본 이미지 기준으로 분할하여 5-fold cross-validation을 수행하였고, 실험은 1,000 회 반복하여 평균 성능을 보고하였다.

III. 전체 배율에서의 성능 비교

표 1은 SREB 데이터셋의 전체 배율($\times 2$, $\times 4$)에 대해 IQA 지표들의 성능을 비교한 결과를 나타낸다. 전반적으로 NR-IQA 지표가 RR-IQA 지표보다 높은 상관계수와 낮은 오차를 보였다. NR-IQA 중 KLTSRQA는 PLCC 0.900, SRCC 0.861로 가장 우수한 성능을 기록하였다. 반면 BRISQUE, SRmetric, ARNIQA는 상대적으로 안정적인 성능을 보였으나, 초해상화로 인한 품질 향상 효과를 충분히 반영하지 못하였다. RR-IQA 지표는 전반적으로 낮은 성능을 보여, 저해상도 참조 정보만으로는 초해상화 영상의 품질 변화를 설명하는 데 한계가 있음을 확인하였다.

표 1. SREB 전체 배율에서의 성능 비교 결과

Method		PLCC	SRCC	RMSE
NR-IQA	BRISQUE	0.865	0.826	0.781
	SRmetric	0.875	0.817	0.757
	KLTSRQA	0.900	0.861	0.680
	ARNIQA	0.883	0.850	0.730
RR-IQA	PSCT	0.373	0.339	1.463
	PFIQA	<u>0.893</u>	<u>0.854</u>	<u>0.700</u>

표 2. SREB $\times 2$ 배율에서의 성능 비교 결과

Method		PLCC	SRCC	RMSE
NR-IQA	BRISQUE	0.880	0.809	0.771
	SRmetric	0.870	0.799	0.804
	KLTSRQA	0.915	0.859	0.655
	ARNIQA	<u>0.904</u>	<u>0.835</u>	<u>0.701</u>
RR-IQA	PSCT	0.505	0.462	1.448
	PFIQA	0.891	0.829	0.742

표 3. SREB $\times 4$ 배율에서의 성능 비교 결과

Method		PLCC	SRCC	RMSE
NR-IQA	BRISQUE	0.874	0.720	0.638
	SRmetric	0.860	0.712	0.667
	KLTSRQA	0.890	0.730	0.600
	ARNIQA	<u>0.886</u>	0.743	<u>0.607</u>
RR-IQA	PSCT	0.458	0.407	1.176
	PFIQA	0.874	0.743	0.640

IV. 단일 배율에서의 성능 비교

표 2 와 표 3 은 각각 SREB 데이터셋의 $\times 2$ 배율과 $\times 4$ 배율에 대해 IQA 지표들의 성능을 비교한 결과를 나타낸다. $\times 2$ 및 $\times 4$ 배율 모두에서 NR-IQA 지표가 RR-IQA 지표보다 우수한 성능을 보였으나, 배율 증가에 따라 대부분의 지표에서 성능 저하가 관찰되었다. KLTSRQA 는 두 배율 모두에서 가장 안정적인 성능을 유지하였으나, $\times 4$ 배율에서는 순위 예측 성능인 SRCC 가 크게 감소하였다. 이는 배율 증가에 따라 발생하는 구조적 왜곡과 인위적 질감 변화가 화질 평가를 어렵게 만들기 때문으로 해석된다.

V. 결론

본 논문에서는 SREB 데이터셋을 기반으로 초해상화 영상에 대한 최신 IQA 지표들의 성능을 비교·분석하였다. 실험 결과, 기존 IQA 지표들은 초해상화 영상의 품질 향상과 왜곡이 공존하는 특성을 충분히 반영하지 못하였으며, 배율 증가에 따라 성능 저하가 두드러졌다. 이는 향후 초해상화 영상 화질 평가를 위해 품질 향상 효과와 배율 의존성을 함께 고려하는 새로운 IQA 접근법이 필요함을 시사한다.

참 고 문 헌

- [1] Y. Kim, J. Shin, J. Lee, and H. Ko, "Subjective and objective quality evaluation of super-resolution enhanced broadcast images on a novel sr-iqa dataset," IEEE Access, vol. 13, pp. 168 087– 168 102, 2025.
- [2] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," IEEE

Transactions on Image Processing, vol. 21, no. 12, pp. 4695– 4708, 2012.

- [3] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," Computer Vision and Image Understanding, vol. 158, pp. 1– 16, 2017.
- [4] Q. Jiang, Z. Liu, K. Gu, F. Shao, X. Zhang, H. Liu, and W. Lin, "Single image super-resolution quality assessment: A real-world dataset, subjective studies, and an objective metric," IEEE Transactions on Image Processing, vol. 31, pp. 2279– 2294, 2022.
- [5] L. Agnolucci, L. Galteri, M. Bertini, and A. Del Bimbo, "ARNIQA: Learning distortion manifold for image quality assessment," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024, pp. 189– 198.
- [6] K. Zhang, T. Zhao, W. Chen, Y. Niu, J. Hu, and W. Lin, "Perception-driven similarity-clarity tradeoff for image super-resolution quality assessment," IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 7, pp. 5897– 5907, 2024.
- [7] X. Lin, X. Liu, H. Yang, X. He, and H. Chen, "Perception- and fidelity-aware reduced-reference super-resolution image quality assessment," IEEE Transactions on Broadcasting, vol. 71, no. 1, pp. 323– 333, 2025.