

# 3차원 웨어렛 변환과 심층 잔류 신경망을 이용한 무참조 스포츠 비디오 화질 평가

이기용<sup>†</sup>, 신승수<sup>\*\*</sup>, 김형국<sup>\*\*\*</sup>

## No-Reference Sports Video-Quality Assessment Using 3D Shearlet Transform and Deep Residual Neural Network

Gi Yong Lee<sup>†</sup>, Seung-Su Shin<sup>\*\*</sup>, Hyoung-Gook Kim<sup>\*\*\*</sup>

### ABSTRACT

In this paper, we propose a method for no-reference quality assessment of sports videos using 3D shearlet transform and deep residual neural networks. In the proposed method, 3D shearlet transform-based spatiotemporal features are extracted from the overlapped video blocks and applied to logistic regression concatenated with a deep residual neural network based on a conditional video block-wise constraint to learn the spatiotemporal correlation and predict the quality score. Our evaluation reveals that the proposed method predicts the video quality with higher accuracy than the conventional no-reference video quality assessment methods.

**Key words:** No-Reference Video Quality Assessment, 3D Shearlet Transform, Natural Scene Statistics, Deep Residual Neural Network, Conditional Constraints

### 1. 서 론

인간이 참조 이미지 없이 왜곡된 이미지의 품질을 평가할 수 있다는 것은 잘 알려진 사실이다. 이미지의 품질 평가는 인간 관찰자의 뇌에 저장된 자연 이미지에 대한 지식과 왜곡된 이미지 사이의 자연스러운 비교를 통해 이루어진다. 이상적이며 객관적인 화질 평가방법은 이러한 인간 관찰자의 품질 예측을 모방할 수 있어야 한다. 객관적 품질 평가방법 중 하나인 NR-VQA(no-reference video quality assessment, 무참조 비디오 화질 평가)는 열화 된 영상만을 가지고 화질을 평가하는 방법으로, 화질의 기준이 되는 원본 영상 혹은 참조 영상에 대한 정보가 전혀

필요하지 않다. 이러한 장점으로 인해 NR-VQA 방식은 실시간 화질평가 및 자동화된 품질관리와 같은 다양한 응용분야에 적용될 수 있으며, 최근 들어 이에 대한 연구가 많은 관심을 받고 있다.

왜곡이 없이 깨끗한 자연스러운 영상에서 움직임은 시간이나 공간에 따라 임의로 변경되지 않으며 실질적인 시공간 상관관계를 가진다. 이러한 시공간상의 통계적 일관성을 NSS(natural scene statics, 자연 장면 통계)[1]라 하며 비디오에 왜곡이 유입될 경우 시공간 상관관계가 흐트러지게 된다. 이로 인해 NSS와의 편차가 발생하고 이러한 편차를 비디오의 화질 평가 지표로 적용할 수 있다. 최근에는 비디오 품질 평가를 회귀문제로 해석하고 다양한 기계학습

\* Corresponding Author : Hyoung-Gook Kim, Address: KwangWoon University, 20 Gwangun-ro, Nowon-gu, Seoul 01897, Republic of Korea, TEL : +82-2-940-5574, FAX : +82-2-913-5006, E-mail : hkim@kw.ac.kr  
Receipt date : Jun. 1, 2020, Revision date : Jul. 21, 2020  
Approval date : Aug. 10, 2020

<sup>†</sup> Dept of Electronics Convergence, Kwangwoon University (E-mail : agayong93@hanmail.net)

<sup>\*\*</sup> Dept of Electronics Convergence, Kwangwoon University (E-mail : tlstn114@naver.com)

<sup>\*\*\*</sup> Dept of Electronics Convergence, Kwangwoon University (E-mail : hkim@kw.ac.kr)

과 딥러닝 방식을 사용하여 비디오 시퀀스의 시공간적 상관관계를 직접 학습함으로써 NR-VQA의 성능을 향상시키는 연구들이 진행되고 있다. J. Søgaard 외[2]는 DCT(discrete cosine transform, 이산 코사인 변환) 계수와 I-프레임의 양자화 레벨 정보를 통해 추출된 공간적 특징과 코덱 특징의 시간 풀링을 통해 획득된 시간적 특징을 서포트 벡터 회귀 방식에 적용하여 H.264/AVC 및 MPEG-2에 의해 왜곡된 비디오의 품질을 평가하였다. 그리고 K. Zhu 외[3]는 DCT 계수로부터 프레임 단위의 특징을 추출하여 자연장면의 왜곡을 정량화하고, 각 프레임 단위의 특징을 시간 풀링을 통해 비디오 단위의 특징으로 변환시켜 다중 신경망에 적용하였으며, Wang 외[4]은 Alex Net을 통해 비디오 프레임 레벨에서 추출한 공간 특징과 비디오의 움직임 벡터에 대한 표준 편차를 통해 추출한 시간적 특징을 다중 회귀방식에 적용하여 화질을 평가하였다. 특히, Y. Li [5]는 3DST(3-dimensional shearlet transform, 3차원 셰어렛 변환)을 이용하여 비디오 블록에서 추출한 평균 시공간 특징에 대한 통계적 회귀모델을 훈련시키기 위해 CNN(convolutional neural network, 합성곱 신경망)을 적용하여 평가 성능을 향상시키는 효율적인 방법을 제안했다. 하지만 이 방식에는 네트워크 상태의 변동으로 인해 발생하는 비디오의 왜곡이 모든 프레임에서 동일하지 않을 수 있다는 사실이 반영되지 않았다. 보다 개선된 품질 평가를 위해서는 회귀학습을 위해 주어진 화질 레이블로부터 프레임 또는 블록 단위의 화질을 예측하여 전체 비디오의 화질 평가에 반영할 수 있는 방법이 필요하다.

본 논문에서는 3DST와 ResNet(deep residual neural network, 심층 잔류 신경망)을 이용한 스포츠

비디오의 NR-VQA방법을 제안한다. 제안된 방식의 주요 관점은 다음과 같다. 1) 블록 단위의 화질 평가를 위해 비디오 시퀀스를 일정한 개수의 프레임들로 이루어진 비디오 블록으로 분할하고 3DST 기반의 시공간 특징을 추출한다. 2) 비디오 시퀀스의 시공간 상관관계를 세밀하게 학습하여 화질 예측 정확도를 향상시키기 위해 ResNet과 로지스틱 회귀를 결합하여 사용한다. 3) 목적함수에 CBC(constraint block-wise condition, 조건부 블록별 제약) 방법[6]을 적용하여 블록 단위의 화질을 자동으로 학습하고 높은 성능으로 비디오의 화질을 평가한다. 본 논문의 2장에서는 제안하는 방법을 설명하고, 3장에서는 실험 결과를 제시하며, 4장에서 결론을 맺는다.

## 2. 제안하는 무참조 비디오 화질 평가 방법

제안하는 NR-VQA의 전체 구조도는 Fig. 1과 같으며, 화질 학습 단계와 화질 추정 단계로 구성된다. 먼저, 학습 단계에서는 스포츠 비디오 데이터베이스로부터 학습 비디오 신호와 비디오에 대한 주관적 화질 점수가 입력된다. 측정된 주관적 화질 점수는 시공간 특징과 주관적 화질 점수 사이의 매핑 관계를 학습하기 위한 해당 비디오의 정답 화질 점수로 사용된다. 다음 단계로, 입력된 비디오의 프레임을 회색조로 변환하고 비디오 화질 분석에 적합한 크기로 조정하기 위해 프레임의 크기를 624×360으로 변환한 뒤 중앙의 416×240 프레임을 추출한다. 그리고 변환된 비디오를 중첩된 비디오 블록으로 분할하고 3DST 기반의 시공간 특징을 추출한다. 추출된 시공간 특징들과 정답 화질 점수는 ResNet-20과 로지스틱 회귀로 연결된 회귀 모델을 학습하기 위해 사용된다. 이 때, 목적함수에 CBC방법을 적용하여 비디오

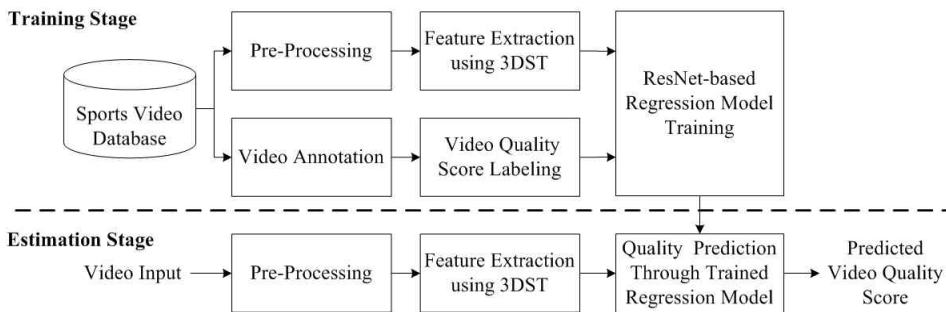


Fig. 1. Flowchart of the proposed NR-VQA system.

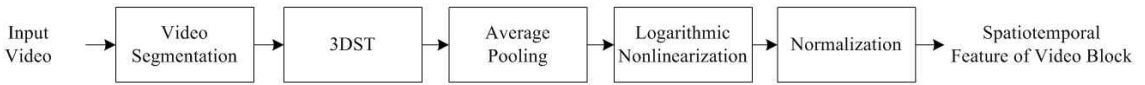


Fig. 2. Framework of spatiotemporal feature extraction from video block.

블록 단위의 화질 점수를 자동적으로 학습한다. 화질 추정 단계에서는 테스트 비디오로부터 추출된 시공간 특징들이 학습된 ResNet-20과 로지스틱 회귀에 입력되어 NR-VQA가 수행된다.

2.1 3차원 웨어렛 변환을 이용한 시공간 특징 추출

본 논문에서는 스포츠 비디오로부터 NSS를 포함하는 시공간 특징을 추출하기 위해 3DST[5]를 적용한다. 3DST는 wavelet[7]변환을 보완하는 다차원 데이터 분석 방법으로 wavelet변환은 주파수영역에 따라 다른 해상도를 적용하여 세부적인 윤곽정보를 효과적으로 검출할 수 있지만 방향정보를 검출할 수 없다. 이와 달리 3DST는 다중 스케일과 다중 방향 분석을 통해 윤곽정보 뿐만 아니라 이방성 특징정보를 검출하기 때문에 다차원 데이터로부터 시공간 특징을 추출하는데 뛰어나며 역동적인 움직임들로 구성된 스포츠 비디오의 시공간 특징을 추출하는데 적합하다.

Fig. 2는 입력된 비디오로부터 시공간 특징을 추출하는 과정을 나타낸다. 먼저, 비디오 블록 단위로 화질을 평가하기 위해 시간 축에 따른 50%의 오버랩을 적용하여 416×240×80(가로×세로×시간)크기의 비디오 블록들로 분할한다. 이 때, 비디오 블록의 크기는 시공간 영역을 훼손하지 않는 범위에서 다양한 크기의 블록을 구성하여 진행한 실험을 통해 최적의 크기를 선택하였다.

다음으로, 분할된 각 비디오 블록은 3DST에 입력된다. 3DST는 다양한 스케일, 방향, 그리고 전단 행렬로 구성된 웨어렛 필터들을 통해 시간 영역에서 다양한 위치와 방향에 따른 wavelet변환을 수행하며 공간 영역에서 모든 주파수를 포함하는 방향 필터의 역할을 한다. 이를 위해 3DST는 이산 푸리에 변환을 통해 입력된 비디오 블록을 주파수 공간으로 변환한다. 비디오 블록의 주파수 공간은 중심부가 원점으로 전체 윤곽 정보를 포함하는 저주파 영역을 나타내고 바깥으로 갈수록 세부 윤곽 정보를 포함하는 고주파 영역을 나타낸다. 그리고 주파수 공간에 대역 통과

필터를 적용하여 세부 윤곽 정보를 포함한 고주파 영역을 추출한 뒤 각 직교축을 포함하는 세 개의 피라미드 영역으로 분할한다. 세 개의 피라미드 영역은 Fig. 3과 같다.

세 개의 피라미드 영역은 모든 고주파 영역을 포함하는 라플라시안 피라미드 구조로 여러 스케일을 통한 단계적인 분석을 가능하게 한다. 또한, 각 피라미드 구조를 통해 방향 성분의 범위를 제한할 수 있으며 촘촘한 방향 필터를 구성할 수 있다. 분할된 각 피라미드 영역과 관련된 웨어렛 필터에 의해 구성되는 3DST는 다음과 같이 정의된다.

$$SH(P_1, P_2, P_3; \alpha, c) = P_1(p_1; \alpha, c) \cup P_2(p_2; \alpha, c) \cup P_3(p_3; \alpha, c), \tag{1}$$

여기서,  $P_1, P_2, P_3$ 는 세 개의 피라미드 영역이며,  $\alpha, c$ 는 각각 이방성 파라미터와 위치 요소이다. 또한,  $p_1, p_2, p_3$ 는 세 개의 피라미드 영역으로부터 정의된 웨어렛들을 나타낸다.

3차원 웨어렛 필터는 라플라시안 피라미드 알고리즘을 통해 피라미드 영역을 다중 스케일 성분으로 분해하고 각 스케일 영역을 방향 성분으로 분석한다. 각 방향 성분은 피라미드의 한 쪽에서 피라미드의 중심을 지나 반대쪽으로 진행한다. 본 논문에서는 각 피라미드 영역을 4개의 스케일로 분해하고 각 스케일 영역에 9개의 방향 성분을 적용하였다. 이 때, 두 번째와 세 번째 피라미드 영역에서 첫 번째 피라미드와 인접하여 방향 성분이 중복되는 영역을 제외하고 52개의 3차원 웨어렛 필터들을 구성하였다. 비디오 블록의 주파수 공간은 각 피라미드 영역의 웨어렛 필터와 요소 곱 연산이 수행되고, 역 3DST 변환이 적용되어 비디오 블록과 동일한 크기의 3차원 웨어

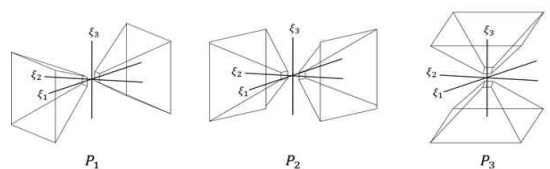


Fig. 3. The partition of frequency domain.

렛 계수로 출력된다. 따라서 하나의 비디오 블록으로부터 52개의 3차원 웨어렛 계수들이 출력된다.

그 후에, 시간과 공간의 변화에 따른 세부적인 상관관계를 분석하기 위해 각 3차원 웨어렛 계수들을 가로, 세로, 시간 축을 따라 2등분하여 8개의 영역으로 분할하고 웨어렛 계수의 차원을 줄이기 위해 각 영역(208×120×40)에 평균 풀링을 적용한다. 풀링 영역 내의 값들은 평균 연산을 통해 단일 값으로 계산되고 하나의 3차원 웨어렛 계수로부터 8개의 풀링 값이 출력된다. 따라서 52개의 3차원 웨어렛 계수로부터 416개의 값이 출력되고 출력된 값은 1차원 벡터로 연결된다. 마지막으로, 벡터의 모든 요소에는 로그 비선형이 적용되고 정규화 되어 시공간 특징으로 출력된다.

2.2 심층 전류 신경망 기반의 회귀 모델

제안하는 회귀 기반의 NR-VQA 방식은 원본 비디오 없이 수신된 비디오만을 가지고 비디오의 화질 점수를 예측하는 것을 목표로 한다. 이를 위해 회귀 모델은 비디오의 시공간 특징과 주관적 화질 점수 간의 매핑 관계를 배우고 예측 기반 화질 매트릭을 생성하도록 학습된다. 하지만 비디오에 유입되는 왜곡의 정도는 모든 프레임에서 동일하지 않기 때문에 효과적인 NR-VQA는 비디오로부터 분할된 비디오

블록에 대한 화질 점수를 예측한 후 이를 전체 비디오에 대한 화질 평가에 반영해야 한다. 또한, 회귀 모델을 생성하기 위해 비디오 데이터베이스로부터 주어지는 정답 화질 점수는 비디오에 대한 화질 점수로 분할된 모든 비디오 블록에 동일한 화질 점수를 할당하는 것은 적절하지 않다. 따라서 NR-VQA는 블록 단위 평가를 반영하는 구조로 설계되어야 한다.

본 논문에서는 비디오로부터 분할된 비디오 블록의 시공간 특징을 입력으로 비디오의 시공간 상관관계를 학습하고 비디오의 화질을 예측하기 위해 ResNet-20[8]과 로지스틱 회귀를 결합한 구조를 제안하며 이는 Fig. 4의 (a)에 나타나 있다. 제안된 방식에서는 목적함수에 CBC방법을 적용하여 블록 단위의 화질 점수를 자동으로 학습함으로써 효과적인 NR-VQA가 수행된다.

ResNet-20 기반의 회귀 모델을 학습하는 과정은 다음과 같다. 먼저, 하나의 비디오 블록으로부터 추출된 3DST 기반의 시공간 특징이 정답 화질 점수와 함께 ResNet-20에 입력된다. 제안하는 ResNet-20의 구조는 합성 곱 필터, 3개의 Residual Block, BN (batch normalization, 배치 정규화), ReLU(rectified linear units), Flatten으로 구성된다. Fig. 4의 (b)와 같이 각 Residual Block은 6개의 합성 곱 층 연결로 구성되고 합성 곱 층은 3개의 합성 곱 필터가 병목구

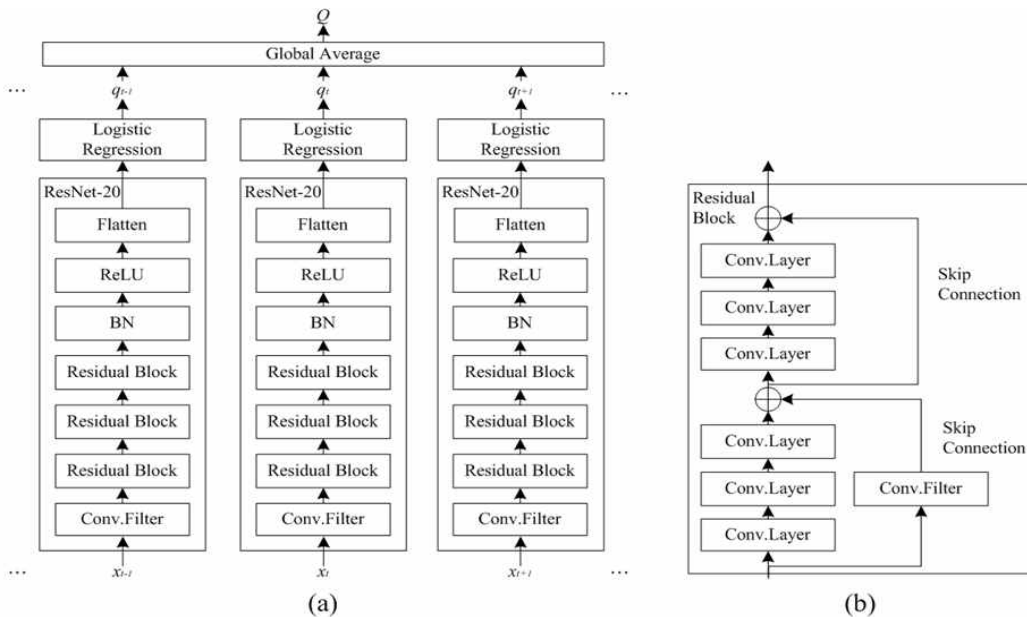


Fig. 4. (a)Block diagram of logistic regression concatenated with ResNet-20 (b)Residual block.

조를 이룬다. 입력된 비디오 블록의 시공간 특징은 ResNet-20의 각 층과 연산된 후 Flatten을 통해 1차원 벡터로 변환되고 로지스틱 회귀를 통해 비디오 블록에 대한 결과 화질 점수로 출력된다. 이렇게 하나의 비디오로부터 분할된 비디오 블록들의 결과 화질 점수들을 모두 계산한 뒤 평균을 구하여 비디오에 대한 최종 결과 화질 점수를 출력한다. 다음으로, 출력 층의 결과 화질 점수들과 최종 결과 화질 점수를 목표 값인 정답 화질 점수와 비교하여 전체 오차를 계산한다. 출력 층은 ResNet-20과 연결된 로지스틱 회귀이며 CBC방법이 통합된 목적 함수의 전체 오차는 다음과 같이 계산된다.

$$O = \frac{1}{S} \sum_{s=1}^S [(\hat{Q}_s - Q_s) + \alpha(\hat{Q}_s) \sum_{t=1}^{T(u_s)} (\hat{Q}_s - q_{s,t})^2], \quad (2)$$

$$\alpha(\hat{Q}_s) = 10^{(\hat{Q}_s - \hat{Q}_{MAX})}, \quad (3)$$

여기서,  $S$ ,  $Q_s$ ,  $T$ 와  $q_{s,t}$ 는 각각 총 학습 비디오의 개수,  $s$  번째 비디오의 최종 결과 화질 점수, 입력 비디오 블록 개수 그리고  $s$ 번째 비디오의  $t$ 번째 블록의 결과 화질 점수이다. 또한,  $\alpha(\hat{Q}_s)$ 는 목적 함수의 가중치이고,  $\hat{Q}_s$ ,  $\hat{Q}_{MAX}$ 는  $s$  번째 비디오의 정답 화질 점수와 최대 화질 점수이다.

식(2)의 첫 번째 부분은 전체 비디오의 화질 예측 정확도에 관련된 부분으로 블록 단위의 화질 분포와는 관련이 없다. 이때, 높은 화질 점수인 비디오의 모든 프레임에는 왜곡이 존재하지 않기 때문에 분할된 각 비디오 블록들은 높은 화질 점수의 균일한 분포를 가져야 하지만 왜곡된 비디오의 경우 특정 프레임에 유입된 왜곡으로 인해 각 비디오 블록들은 균일하지 않는 화질 점수 분포를 가진다. 이를 기반으로 식의 두 번째 부분에서는 식 (3)의 가중치를 적용하여 비디오의 화질 점수에 따라 블록 단위 오차를 전체 오차에 부분적으로 반영한다. 만약, 최대 화질 점수가 5라고 가정할 때, 정답 화질 점수가 5일 경우 가중치는 1로 계산되고 정답 화질 점수가 1일 경우 0.0001의 값을 가지게 된다. 즉, 높은 화질 점수의 비디오로부터 분할된 비디오 블록의 오차는 많이 반영하고 낮은 화질 점수의 비디오로부터 분할된 비디오 블록의 오차는 적게 반영함으로써 높은 화질의 비디오에 집중하여 학습하고 블록 단위 화질이 전체 비디오의 균일한 화질 분포를 따르도록 만든다. 이러한 제약 조건은 ResNet-20이 왜곡이 없는 비디오 블록

과 왜곡이 많은 비디오 블록을 구별하도록 유도한다. 그리고 계산된 전체 오차는 역전과 과정을 통해 출력 층에서 입력 층으로 전달되고 출력 층과 은닉 층 사이의 가중치를 갱신한다. 마지막으로 전체 오차가 지정된 범위 내로 수렴하면 학습을 종료하고 갱신된 가중치를 저장한다.

제안하는 방식은 학습을 위해 주어진 화질 점수로부터 CBC방법을 통해 블록 단위의 화질 점수를 자동적으로 학습함으로써, 전체 비디오에 대한 화질 예측 성능을 향상시킨다. 또한, ResNet-20은 입력과 출력의 차이인 잔차를 학습하는 구조로 입력 값의 작은 변화에도 민감하기 때문에 입력된 비디오 블록의 시공간 특징으로부터 비디오의 시공간 상관관계를 세밀하게 학습한다.

### 3. 실험 결과

#### 3.1 데이터베이스와 실험 방법

제안하는 방법과 다른 NR-VQA 방법들의 성능을 비교하기 위해 두 종류의 비디오 데이터베이스를 사용하였다. 첫 번째 데이터베이스는 스포츠 비디오 데이터베이스로서 60개의 참조 비디오와 480개의 왜곡된 비디오로 구성되어 있으며, 모든 비디오는 960×540 해상도의 YUV420 형식을 사용한다. 왜곡의 종류로는 H.264 손실 비디오 압축 코덱을 사용한 압축 왜곡과 네트워크 트래픽 에뮬레이터를 사용한 패킷 손실을 적용하였다. 각 동영상의 주관적인 화질 평가 점수는 18세에서 40세 사이의 15명의 참가자를 대상으로 MOS(mean opinion score, 평균 의견 점수) 기법에 의해 측정되었다. MOS점수는 5(우수), 4(양호), 3(보통), 2(불량), 1(나쁨) 등의 5단계 등급으로 평가된다. 두 번째 데이터베이스는 스포츠 경기와 사람들의 움직임이 포함된 CSIQ(computational and subjective image quality)[9] 공개 데이터베이스로 12개의 참조 비디오와 216개의 왜곡된 비디오로 구성되며, 모든 비디오는 832×480 해상도의 YUV420 형식이고, 각 비디오의 길이는 10초이다. 또한, 참조 비디오에는 6가지 왜곡 종류(H.264 / AVC 압축, 패킷 손실, 모션 JPEG 압축, Snow 코덱을 사용한 Wavelet 압축, White Noise 및 HEVC 압축)가 적용되었고, 1에서 100 사이의 DMOS(differential mean opinion score, 차동 평균 의견 점수)와 각 비디오에 대한

Table 1. Results of experiments

Database	Method	LCC	SROCC
Sports Video Database	CF-SVR[2]	0.7753	0.7845
	VF-NN[3]	0.7826	0.7937
	COME[4]	0.7965	0.8072
	3dshCNN[5]	0.8031	0.8158
	3dshCNN(CBC)	0.8167	0.8254
	3dshResNet	0.8312	0.8462
	3dshResNet(CBC)	0.8537	0.8654
CSIQ Video Database	3dshCNN	0.8557	0.8524
	3dshCNN(CBC)	0.8645	0.8731
	3dshResNet	0.8691	0.8786
	3dshResNet(CBC)	0.8771	0.8856

DMOS의 표준 편차가 제공되어 있다. 제안하는 NR-VQA의 성능 평가 지표로 LCC (linear correlation coefficient)와 SROCC (spearman rank-order correlation coefficient)를 측정하여 비교하였다[10].

### 3.2 실험 결과

Table 1에서는 스포츠 데이터베이스와 CSIQ 데이터베이스에 대해 제안하는 방식과 다른 신경망 구조 및 특징 값을 사용하는 NR-VQA 방식들을 적용한 결과를 비교한다.

스포츠 데이터베이스에서는 제안하는 방식인 3dshResNet(CBC)가 LCC 0.8537 및 SROCC 0.8654로 가장 우수한 결과를 나타낸다. 3dshResNet(CBC)는 3dshResNet 보다 뛰어나고, 3dshCNN(CBC)는 3dshCNN보다 뛰어난 성능을 보여준다. 이러한 결과를 통해 CBC방법을 적용함으로써 비디오의 화질 평가 성능이 향상됨을 확인할 수 있다. 또한, 3DST를 통해 블록단위로 추출된 시공간 특징을 딥러닝 기반의 회귀예측에 적용한 방식들이 공간적 특징과 시간적 특징을 독립적으로 추출하여 기계학습 방식에 적용한 CF-SVR[2], VF-NN[3] 그리고 COME 방식들 [4] 보다 성능이 더 우수한 것을 확인할 수 있었다.

CSIQ 데이터베이스의 실험 결과에서는 LCC 및 SROCC 측면 모두에서 3dshResNet(CBC)가 가장 높은 성능을 보여주었다. 3dshResNet은 3dShCNN 및 3dShCNN(CBC)보다 나은 결과를 나타내지만 제안하는 방법인 3dshResNet(CBC)보다 낮은 성능을 나타내었다. 이는 동일한 3DST 기반의 시공간 특징

추출방식을 적용하였더라도 CNN 보다도 ResNet이 입력된 비디오 블록의 시공간 특징을 세밀하게 학습함으로써 보다 나은 예측 성능을 제공하는 것을 알 수 있다.

### 4. 결 론

본 논문에서는 3DST를 이용한 시공간 특징을 CBC 기반의 ResNet과 로지스틱 회귀의 결합에 적용한 NR-VQA방법을 제안했다. 두 종류의 데이터베이스에 대한 성능 평가 실험을 통해 제안하는 방법이 가장 높은 정확도로 비디오의 화질을 평가한다는 것을 확인할 수 있었다. 향후에는 사람의 뇌파와 연계하여 시공간 상관관계를 더 효과적으로 학습하는 방법에 대한 연구를 수행할 예정이다.

### REFERENCE

- [1] T.R. Goodall, A.C. Bovik, and N.G. Paulter, "Tasking on Natural Statistics of Infrared Images," *IEEE Transaction Image Process*, Vol. 25, No. 1, pp. 65-79, 2016.
- [2] J. Sogaard, S. Forchhammer, and J. Korhonen, "No-reference Video Quality Assessment Using Codec Analysis," *IEEE Transaction Circuits System Video Technology*, Vol. 25, No. 10, pp. 1637-1650, 2015.
- [3] K. Zhu, C. Li, V. Asari, and D. Saupe, "No-reference Video Quality Assessment Based on

Artifact Measurement and Statistical Analysis,” *IEEE Transaction Circuits System Video Technology*, Vol. 25, No. 4, pp. 533-546, 2014.

[4] Wang, Chunfeng, Li Su, and W. Zhang, “COME for No-reference Video Quality Assessment,” *Proceeding of IEEE Conference on Multimedia Information Processing and Retrieval*, pp. 232-237, 2018.

[5] Y. Li, L.M. Po, X. Xu, L. Feng, F. Yuan, and K.W. Cheung, “No-reference Video Quality Assessment with 3D Shearlet Transform and Convolutional Neural Networks,” *IEEE Transaction Circuits System Video Technology*, Vol. 26, No. 6, pp. 1044-1057, 2016.

[6] S.W. Fu, Y. Tsao, H.T. Hwang, and H.M. Wang, “Quality-Net: An End-to-end Non-intrusive Speech Quality Assessment Model Based on BLSTM,” *Proceeding of Interspeech*, pp. 1873-1877, 2018.

[7] W.D. Chang and C.H. Im, “EOG-based User-independent Gaze Recognition Using Wavelet Coefficients and Dynamic Positional Warping,” *Journal of Korea Multimedia Society*, Vol. 21, No. 9, pp. 1119-1130, 2018.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[9] P.V. Vu and D.M. Chandler, “ViS3: An Algorithm for Video Quality Assessment via Analysis of Spatial and Spatiotemporal Slices,” *Journal of Electronic Imaging*, Vol. 23, No. 1, p. 013016, 2014.

[10] K. Gu, G. Zhai, X. Yang, and W. Zhang, “Hybrid No-reference Quality Metric for Singly and Multiply Distorted Images,” *IEEE Transactions on Broadcasting*, Vol. 60, No. 3, pp. 555-567, 2014.



이 기 용

2020년 2월 광운대학교 전자융합공학과 학사  
2020년 3월~현재 광운대학교 전자융합공학과 석사과정



신 승 수

2019년 2월 광운대학교 전자융합공학과 학사  
2019년 3월~현재 광운대학교 전자융합공학과 석박과정



김 형 국

1999년~2002년 독일 SIEMENS/Cortologic AG 책임연구원  
2002년~2005년 독일 베를린 공과대학교 Assistant Professor  
2005년~2007년 삼성종합기술원 수석 연구원  
2007년 3월~현재 광운대학교 전자융합공학과 교수