

특집논문 (Special Paper)

방송공학회논문지 제23권 제6호, 2018년 11월 (JBE Vol. 23, No. 6, November 2018)

<https://doi.org/10.5909/JBE.2018.23.6.768>

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

## MSSSIM 및 쿨백-라이블러 발산 기반 의사 율-왜곡 평가 함수와 복수개의 영상처리 필터를 이용한 동영상 전처리 방법

석진욱<sup>a)†</sup>, 조승현<sup>a)</sup>, 김휘용<sup>a)</sup>, 최진수<sup>a)</sup>

### Image Processing of Pseudo-rate-distortion Function Based on MSSSIM and KL-Divergence, Using Multiple Video Processing Filters for Video Compression

Jinwuk Seok<sup>a)†</sup>, Seunghyun Cho<sup>a)</sup>, Hui Yong Kim<sup>a)</sup>, and Jin Soo Choi<sup>a)</sup>

#### 요 약

본 논문에서는 동영상 화질을 최대한 유지하면서 압축 비트량 절감을 효율적으로 이루기 위해 복수개의 영상처리 필터를 영상의 블록에 따라 선택적으로 적용하고, 영상처리 필터의 선택을 위한 MSSSIM(Multi-Scale Structural SIMilarity) 및 쿨백-라이블러 발산(Kullback - Leibler divergence: KL-Divergence) 기반의 영상 처리 평가 함수를 제안한다. 영상압축의 경우, 영상 내 특징에 따라 화질과 비트량 절감의 특성이 다르며, 이에 따라 단일 목적을 가진 영상처리 필터로서는 화질을 유지하면서, 비트량 절감이라는 목적을 동시에 만족 시키기 어렵다. 이에 따라, 주관적 화질을 최대한 유지하면서, 비트량을 절감시키기 위해 주관적 화질 측도로서 MSSSIM를 사용하고 비트 량 측도를 위하여 쿨백-라이블러 발산을 사용함과 동시에 두 가지 척도를 하나의 척도로 결합시키기 위한 방법을 제안한다. 아울러 제안한 척도를 사용하여 서로 다른 특성을 가진 영상처리 필터를 전처리 필터로 사용할 경우, 주관적 화질을 최대한 유지하면서 비트량 절감을 유지할 수 있도록 동영상 압축이 가능함을 확인할 수 있었다.

#### Abstract

In this paper, we propose a novel video quality function for video processing based on MSSSIM to select an appropriate video processing filter and to accommodate multiple processing filters to each pixel block in a picture frame by a mathematical selection law so as to maintain video quality and to reduce the bitrate of compressed video. In viewpoint of video compression, since the properties of video quality and bitrate is different for each picture of video frames and for each areas in the same frame, it is difficult for the video filter with single property to satisfy the object of increasing video quality and decreasing bitrate. Consequently, to maintain the subjective video quality in spite of decreasing bitrate, we propose the methodology about the MSSSIM as the measure of subjective video quality, the KL-Divergence as the measure of bitrate, and the combination method of those two measurements. Moreover, using the proposed combinatorial measurement, when we use the multiple image filters with mutually different properties as a pre-processing filter for video, we can verify that it is possible to compress video with maintaining the video quality under decreasing the bitrate, as possible.

Keyword : MSSSIM, KL-Divergence, Video Filtering, pseudo-rate-distortion

Copyright © 2016 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

## 1. 서론

최근 들어 통신 기술의 급격한 발전에 힘입어, 음성 신호와 영상 신호를 포함하는 멀티미디어 데이터의 이용이 급속도로 증가하고 있다. 그러므로 동영상에 대한 수요도 지속적으로 증가하고 있으며 따라서 효과적인 동영상의 화질 향상 기법이 요구되고 있다. 특히 기계학습 알고리즘의 발전으로 인하여 정지영상의 손실된 영상이나, 존재하지 않는 영상을 기계학습 기반의 알고리즘을 통해 사용자가 원하는 영상의 통계적 특성을 기반으로 하는 평가함수를 최적화 시키는 방식으로 얻어내는 연구가 성공적으로 진행되고 있다<sup>[1][2]</sup>. 하지만, 상기 서술한 기존 기계학습 방법론들은 대체로 정지영상을 대상으로 영상의 변형, 축소, 확대 및 왜곡에 강한 영상을 만들어내는 것에 초점을 맞추고 있어 동영상 압축에서 연구 대상으로 하는 비트 율 질감 및 화질개선을 위한 기술 개발이라는 측면에서 충분한 연구가 이루어지지 않은 상황이다.

본 논문에서는 기존 알고리즘을 기반으로 동영상 압축 시 주관적 화질을 최대한 유지하면서, 상대적으로 비트 율을 절감하기 위한 영상 전처리 방법론을 제시한다. 일반적으로 동영상 압축 비트 율을 줄이기 위해서는 동영상 압축 코덱 내부의, 이산 여현 변환(Discrete Cosine Transform: 이산 여현 변환)과 양자화(Quantization)과정에서 최대한 발생 정보가 최소화 될 수 있도록 하여야 한다. 이를 위해서는 여러 가지 방법론이 제시되고 있으나 보통의 경우 입력 영상에 매우 근접한 예측 영상의 생성과 인-루프(In Loop Filter)필터로 대표되는 양자화 왜곡 최소화 영상 처리 방법

론 및 압축자체의 알고리즘 최적화를 들 수 있다. 또 다른 방법으로는, 이산 여현 변환 변환 공간에서 주파수 별 계수의 분포가 압축 정보량을 최소화 시킬 수 있는 분포로 나타나거나, 양자화 왜곡의 영향을 줄일 수 있는 분포로 만드는 방법이 있을 수 있다. 이러한 방법론의 대표적인 것이 주관적 화질을 고려하여 양자화 계수의 값을 조절하는 최소 가지 차이 (Just Noticeable Difference: JND) 방법론이 있다<sup>[3]</sup>. 최소 가지 차이 방법론은 영상 내 특정 화소를 중심으로 휘도 성분의 평균값과 주변 화소의 휘도 차를 사람이 얼마나 잘 인식 할 수 있는가를 검사하여 잘 인식 할 수 없다면 해당 영상의 양자화 계수 값을 크게 하여 발생 비트량을 조절하는 방법이다. 그런데 이 방법을 영상 자체에만 놓고 보면, 사실상, 매우 효율적인 백색잡음 제거 방법론과 유사하며, 대체로 고주파 성분에 대하여 양자화 계수 값을 조절하게 된다는 측면에서도 영상 내 백색 잡음 제거와 유사하다. 따라서, 만일, 특성이 서로 다른 영상 내 백색잡음 제거 필터를 여러 개 두고, 영상 내부의 특성에 따라 제한된 영역에 대하여 특정 필터를 적용하는 방식으로 영상 전처리를 가하게 된다면, 최소 가지 차이 방법을 통한 효과와 유사하거나, 더 나은 성능을 기대할 수 있을 것이다.

본 논문은 이러한, 접근 방법을 구현하기 위하여 주관적 화질 측도로써 많이 사용되는 MSSSIM<sup>[4]</sup>을 기준으로 비트 율 절감효과가 상이한 3가지 영상 처리 필터를 선택하기 위한 화질 및 비트율 척도를 제안하고, 이에 따른 영상 전처리 결과와 동영상 압축 적용시의 특성을 보인다. 본 논문의 구성은 다음과 같다, 2절에서는 본 논문이 제안하는 방법론에 대한 구체적인 설명을 하고, 3절에서는 영상처리 필터를 선택하기 위한 화질 및 비트 율 척도를 유도한다. 4절에서는 제안한 기법의 성능을 실험을 통해 확인하며 마지막으로 5절에서는 본 논문의 결론을 맺는다.

## II. 영상처리 필터뱅크와 특성

### 1. 의사 율-왜곡 측도의 필요성

최근 심층 신경망 (Deep Neural Network: DNN), 특히 이미지 처리를 위하여 심층 신경망의 각 층 (Layer)를 2차원 구조로 놓고 여기에 간단한 2차원 영상 필터의 결과에

a) 한국전자통신연구원 방송·미디어연구소 미디어연구본부 실감AV연구그룹 (Realistic AV Research Group Media Research Division Broadcasting-Media Research Laboratory Electronics and Telecommunications Research Institute)

‡ Corresponding Author : 석진욱(Jinwuk Seok)

E-mail: jnwseok@etri.re.kr

Tel: +82-42-860-6465

ORCID: <https://orcid.org/0000-0001-5318-1237>

※ 이 논문의 연구결과 중 일부는 “한국방송·미디어공학회 2018년 하계학술대회”에 발표한 바 있음.

※ 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.18HR2300, 초실감 테라미디어를 위한 AV부호화 및 LF미디어 원천기술 개발).

· Manuscript received September 7, 2018; Revised November 8, 2018; Accepted November 8, 2018.

가중치를 두어 심층 신경망의 구조로 구현한 컨볼루션 신경망 (Convolutional Neural Network: CNN) 기반의 영상 처리가 활발히 연구되고 있다<sup>6)[7]</sup>.

그런데, 컨볼루션 신경망기반의 영상처리를 살펴보면, 매우 기본적인 영상필터의 결과에 대하여 서로 다른 가중치를 주고 이것을 다층 레이어 신경망에 의해 가중치를 놓아 학습시킨 결과라는 접근 방향은 맨 마지막, 혹은 출력 레이어에서 얼마 떨어지지 않은 레이어의 결과를 제외하고 나머지 레이어에서의 결과는, 사실상, 일반적인 개념의 영상처리 결과와 동일하다는 점이다. 즉, 컨볼루션 신경망 기반 영상처리에서 거의 대부분의 레이어에서의 학습 결과는 사실상, 영상처리와 동일하며, 은닉 층에 저장된 서로 다른 영상처리 결과를 마지막 출력 층에서 선택하여 출력을 내보내는 것과 같다. 따라서 만일, 기존의 여러 가지 방법의 영상 처리 필터를 놓고, 이 중에서 특정 영역에 대해서 특정한 영상 처리

결과만을 선택하도록 하면, 이는 사실상, 컨볼루션 신경망 기반 영상처리와 거의 동일한 개념이라 볼 수 있다.

이를 위해서는, 기존 컨볼루션 신경망이 영상의 전역적 특성을 표현하는 목적함수를 사용하는데 반해, 목적함수가 특정 영상 특성에 대하여 특정 영상 처리 필터를 선택할 수 있도록 만들어 주고 이에 따라 필터를 선택할 수 있게 만든다면, 비트율을 떨어뜨리면서, 영상의 주관적 품질을 유지시키고자 하는 목표를 쉽게 달성할 수 있을 것이다. 이러한 접근 방법을 선택하게 되다면, 심층신경망을 사용한 영상처리에서 학습 후, 결과만을 사용한다는 개념이, 기존의 영상처리 알고리즘 중 특정 영상의 특정 블록에 잘 적용될 수 있는 알고리즘을 선택하는 문제로 바뀌어지게 된다. 따라서, 비트율과 화질 측도를 각각 계측할 수 있는 함수를 놓고 이를 결합시켜 영상 알고리즘을 선택하도록 하는 의사 율-왜곡 함수를 만들게 되면, 심층 신경망보다

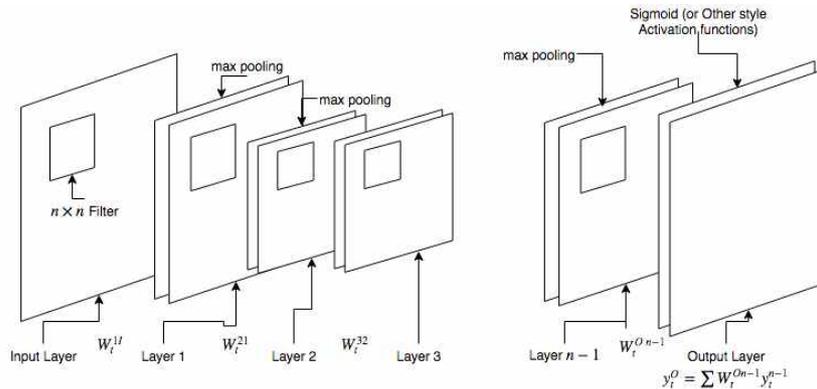


그림 1. 컨볼루션 신경망에 기반한 영상 처리용 심층 신경망 구조  
 Fig. 1. The structure of DNN for image processing based on 컨볼루션 neural network

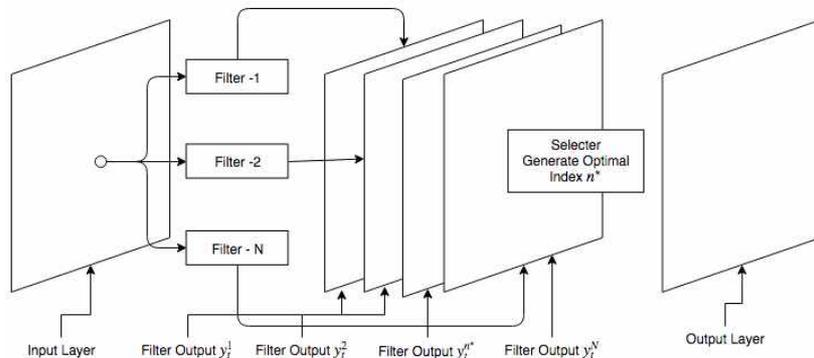


그림 2. 본 논문에서 제안한 방식. 적절한 율-왜곡 평가함수를 통해 여러 영상처리 결과 중 하나를 선택한다.  
 Fig. 2. The proposed structure in this paper. By an appropriate rate-distortion function, select the output from plural outputs of image processing filters

효과적인 영상 처리 방법론을 보다 간편하게 구현할 수 있는 방법론이 될 것이다.

## 2. 의사 율-왜곡 측도의 필요성

서론에서 논한 바와 같이, 인간의 시각 시스템 (HVS: Human Visual System)은 매우 복잡한 구조의 시스템으로, 다양한 특성을 갖지는 바, 그 중 가장 우세한 특성 중 하나는 휘도 적응(luminance adaptation)이다. 보통 휘도 적응 특성의 결과로 매우 어두운 영역에서의 밝기 변화에 대해서는 인간 시각 특성에 매우 둔감한 편이며 밝은 영역에 대해서는 민감하게 나타난다. 그러나, 밝은 영역에서의 민감한 특성은 인간이 영상의 구조적 특성을 잘 표현해 줄 수 있는 특성, 예를 들어 사물의 모서리(Edge)부분 등에 민감한 것이지 영상의 구조적 특성을 잘 묘사할 수 없는 부분, 예를 들어 각종의 잡음 등에 대해서는 오히려 화질이 떨어지는 것으로 받아들여지게 된다. 그러므로, 영상처리 필터는 영상의 구조적 특성을 잘 반영하면서도 구조화 될 수 없는 영상신호는 제거할 수 있어야 하며 이러한 특성을 잘 반영하는 필터의 종류로 웨이블릿 (Wavelet) 기반의 영상처리 필터 혹은 다운샘플링 기반 필터가 많이 사용된다.

웨이블릿 기반 영상처리 필터는 영상을 다단계 웨이블릿 변환을 거친 후 특정 임계 치 (Threshold)값 보다 작은 값의 경우 0으로 처리하고, 3x3 블록의 평균값보다 큰 값 중 일정한 값 이상은 더 크게 확대 시키는 방식으로 임계치의 크기와 블록 평균값과의 차이에 따라 영상처리 결과가 달라진다. 보통 고주파 대역의 값에서 이러한 변화가 많이 나타나게 되므로 자연스럽게 영상의 잡음성분을 제거하게 된다. 또 하나는 다운샘플링 기반 필터로서 영상을 원본 영상의 절반 크기로 줄인 다음 여기에 위너 필터링(Wiener-Filtering)을 적용하고 그 결과를 그대로 원본 영상으로 되돌리는 것으로서, 매우 강력한 잡음 제거 성능을 보이지만, 축소된 영상을 다시 원본 영상으로 되돌리는 과정에서 Blur 왜곡이 발생한다. 본 논문에서는 다운샘플링 기반 필터, 강한 웨이블릿 기반 필터, 약한 웨이블릿 기반 필터의 3가지 필터를 사용하며, 일반적으로 다운샘플링 기반 처리된 영상은 가장 화질이 좋지 않으나 비트량 절감 효과가 가장 우수하고, 강한 웨이블릿 기반 필터는 화질과 비트량 절감 효과가 중간 정도이며, 약한

웨이블릿 기반 필터는 가장 좋은 화질을 제공하나 가장 비트량 절감 효과가 떨어진다. 그러므로, 영상의 국소적인 구조적 특성에 따라 3가지 필터 중 적절한 필터를 선택하여 영상을 전 처리하게 되면 비트량 절감과 주관적 화질 열화의 최소화라는 두 가지 목표를 달성할 수 있다.

## III. MSSSIM과 콜백-라이블러 발산 기반 의사 율-왜곡 측도

MSSSIM은 원본 영상 및 다운 샘플링 된 영상에 대하여 구조적 유사 지수 (Structural Similarity: SSIM)를 구한 뒤 이의 가중 평균을 통해 영상의 주관적 화질을 측정하는 지표로서 최근 구조적 유사 지수에 비하여 인지화질 측정 측면에서 보다 높은 정합 율을 보이는 특성이 있다.

처리된 영상의 비트량 예측의 경우, 일반적으로 실제 영상 압축 후 나타나는 비트량을 그대로 사용할 수 있다면 가장 좋은 방법이 될 수 있겠으나, 전처리 필터의 경우, 이 정보를 그대로 사용하는 것은 영상 코딩을 실제로 수행한 이후에 적용할 전처리 필터를 선택하는 것이기 때문에 비트량을 적절히 예측할 수 있는 함수를 만들어야 한다.

동영상 압축은 기본적으로 이산 여현 변환 변환공간에서 양자화를 수행한 후 (대수적)엔트로피 코딩을 수행하는 것이기 때문에 양자화 결과를 생각하지 않는다면, 이산 여현 변환 공간에서의 엔트로피를 유추할 수 있어야 한다. 또한, 영상처리를 통해 잡음 성분이 제거되었다고 가정하면, 이에 해당하는 엔트로피가 제거된 것이기 때문에, 원본영상의 이산 여현 변환 공간에서의 엔트로피 보다, 작을 것이기 때문에 원본영상과 처리된 영상간의 엔트로피의 차이가 비트량의 차이와 비례할 것이며 그러므로 원본영상과 처리 영상간 2차원-이산 여현 변환 변환 값의 분포에 대한 콜백-라이블러 발산이 절감되는 비트량으로 유추할 수 있다.

영상의 특정 블록을  $X \in \mathbf{R}^{n \times n}$ 라 놓고 이산 여현 변환 변환 행렬을  $T \in \mathbf{R}^{n \times n}$ 라 하면 2차원-이산 여현 변환 변환은  $TXT^T$ 로 정의되고 이에 대한 고유치  $e$ 는 다음과 같이 구해진다.

$$\arg_{\det}(TXT^T - eI) = 0 \quad (1)$$

이때, 각 고유치의 값을 크기 별로 정렬한 고유치 값의

벡터  $e \in \mathcal{C}^n$ 의 파워 스펙트럼(Power Spectrum)을 다음과 같이 공액 복소수 (Conjugate) 연산 결과로 정의한다.

$$e = \sqrt{e \oplus e^H} \quad (2)$$

이때,  $e \in \mathcal{R}$ 는 다음과 같이 정의한다.

$$e_0 \geq e_1 \geq \dots \geq e_n \text{ such that } e = (e_0, e_1, \dots, e_{n-1})^T \\ \forall k \in \mathcal{Z}(0, n-1], e_k \in \mathcal{R}$$

여기서 고유치 파워 스펙트럼의 k번째 성분  $e_k > 0$ 의 발생 확률을 다음과 같이 정의하고 이렇게 이루어진 분포를 p라고 하자.

$$p_k = \frac{e_k}{\sum_{l=0}^{n-1} e_l}, e_l \in e, \therefore \sum_{k=0}^{n-1} p_k = 1 \quad (3)$$

원본 영상블록(ex: 32x32 크기의 1 프레임 (Frame)내 t번째 block)  $R_t$ 와 i번째 처리된 영상  $\bar{P}_{i,t}$ 가 존재한다고 할 때  $p \sim \bar{P}_{i,t}$ 이고  $q \sim R_t$ 라고 하면, 이 두 분포의 엔트로피 차이는 쿨백 라이블러 발산  $D_{KL}$ 이며 이를 두 영상간의 비트율 추정함수 후보  $\tilde{R}$ 로 다음과 같이 정의한다.

$$\tilde{R} = D_{KL}(p \parallel q) = \sum_k (p_k \log p_k - p_k \log q_k) \quad (4)$$

그런데, 실제로, MSSSIM 값과, 값의 크기 문제가 있어 이를 결합하기 위한 보정함수가 필요하다. 보정함수는 영상의 특성에 따라 먼저, 단순한 영상의 경우 MSSSIM이 처리 방법에 따라 비슷하게 나타나기 때문에 가급적 비트량을 절감할 수 있는 결과를 선택되도록 하여야 하며 이를 위해  $D_{KL}$ 의 값의 비중이 MSSSIM에 비하여 높게 나타나야 하며 복잡한 영상의 경우 MSSSIM이 높게 나타나도록  $D_{KL}$  값의 비중을 줄이도록 한다. 특히, 적당히 복잡한 영상 혹은 단순한 배경 위에 모서리 부분이 두드러지는 경우 최대 신호 대 잡음비 (Peak Signal Noise Ratio: PSNR)을 기준으로 일정 값 이하에서는 MSSSIM의 비중이 상대적으로 높게 나타나도록 한다.

이를 위해 먼저 단순한 영상과 복잡한 영상에 따른  $D_{KL}$ 의 비중 조절을 위해 DC성분값에 반비례하고 그 외 성분에  $D_{KL}$  값이 비례하도록 다음과 같이 놓는다.

$$\frac{1-p_0}{p_0} \cdot D_{KL} \quad (5)$$

식 (5)에서  $p_0$ 는 고유치 파워 스펙트럼의 DC성분 값으로 식 (3)에서 정의된  $p_k$ 에서 가장 큰 파워 스펙트럼  $e_0$ 에 대응하는 분포 값이다. 식 (3)에서  $\sum_{k=0}^{n-1} p_k = 1$ 이므로

$$1-p_0 = \sum_{k=0}^{n-1} p_k - p_0$$

가 되어 식 (5)의 분자 항은 DC성분을 제외한 나머지 성분이 된다. 반면, 분모 항은 DC항이므로 식 (5)를 통해는 단순한 영상에서는 복잡한 영상에서 보다 작은 값을 가지게 되며,  $0 < p_k < 1$ 이므로, 식 (5)의 값은  $D_{KL}$  보다 작은 값을 가지면서, 단순한 영상에서 더 작은 값을 가지도록 한다.

그리고 특정 최대 신호 대 잡음비 이하에서는 신호 대 잡음비의 비중이 크고 비트 량의 추정 함수인  $D_{KL}$  값의 비중이 작게 나타나도록 하여, 비트가 많이 발생하더라도 가급적 최대 신호 대 잡음비의 값이 높게 나타나는 필터가 선택되도록 한다. 반대로 특정 최대 신호 대 잡음비 이상에서는 신호 대 잡음비의 비중을 낮추고  $D_{KL}$ 의 비중을 높여 비트 량을 절감시킬 수 있는 필터가 선택될 수 있도록 한다. 마지막으로 이러한 과정에서 사용되는 함수는 연속 함수로서 정의되도록 한다. 이러한 목적을 달성하기 위해서는 MSSSIM을 신호간 왜곡으로 보고, 비트 량의 추정으로 간주하는 엔트로피의 차이를 잘 정의된 라그랑지안 승수로서 결합시켜야 한다.

그라스마니안(Grassmanian)기반 해석에 의하면, 임의의 라그랑지안 (혹은 헤밀토니안)의 라그랑지안 승수는 시스템 상태 계수로 생성되는 공간의 선형 변환으로 매핑되는 부분 공간내의 원소로 정의된다<sup>[8]</sup>. 따라서, 구하고자 하는 목적함수의 상태 변수는 식 (3)에서와 같이 파워 스펙트럼의 발생 확률로 정의되므로 라그랑지안 승수 역시, 파워 스펙트럼의 확률 밀도 함수의 형태로 정의 되어야 한다. 이때, 파워 스펙트럼의 발생 확률이 페르미-디랙 통계를 따른다고 가정하면 라그랑지안 승수는 시그모이드 함수의 형태로 나타나는 것이 적절하며 다음과 같이 정의한다.

$$\lambda(E_k) = 10.0 \left( \frac{0.99}{1 + \exp(\beta(m - E_k))} + 0.01 \right) \quad (6)$$

식 (6)에서  $\beta$ 는 시그모이드 함수의 기울기 계수로 2를 주었는데, 이는 실험을 통해서 적절한 필터가 선택될 수 있도록 최소 자승법을 통해 1차적으로 구하고 이어, 최소 자승 법으로 구해진 값의 가장 가까운 정수 값을 사용하여

정하였다.  $E_k$ 는 영상  $R_t$ 와  $\bar{P}_{i,t}$ 의 k 번째 블록의 최대 신호 대 잡음비이며  $m$ 은 기준 최대 신호 대 잡음비로서 자유롭게 정할 수 있으며 20~50 사이의 값에서 선택하면 된다. 따라서 식 (6)을 사용하여  $E_k$ 가 기준 최대 신호 대 잡음비

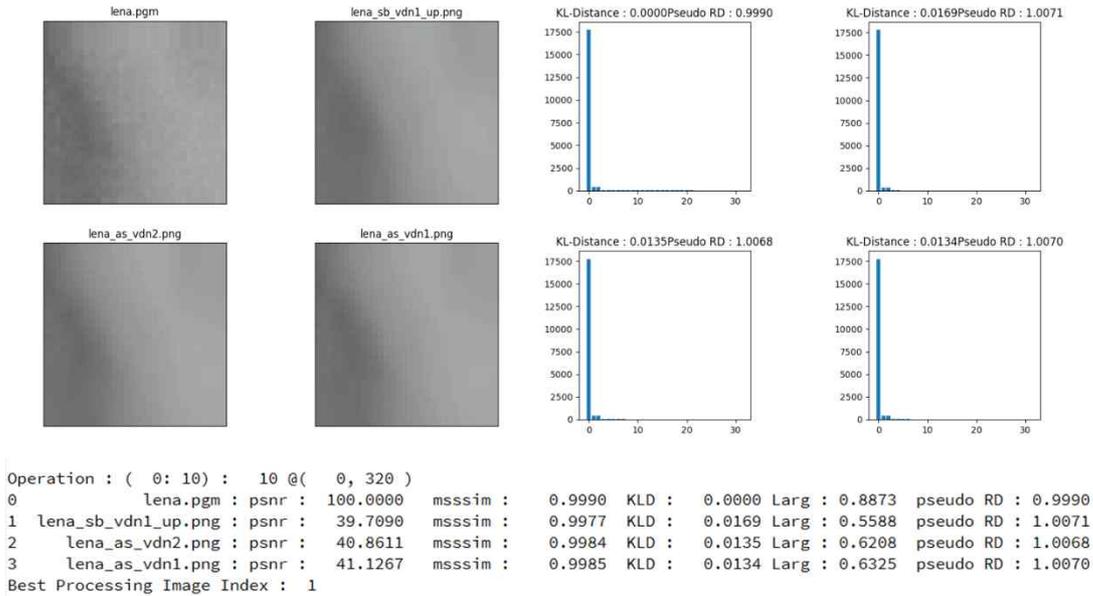


그림 3. 영상처리 필터 1번(sb\_vdn1\_up)이 선택되는 경우의 제안한 의사 율-왜곡 함수의 특성  
 Fig. 3. The characteristics of the proposed pseudo rate-distortion function when the image processing filter No.1 (sb\_vdn1\_up) is selected

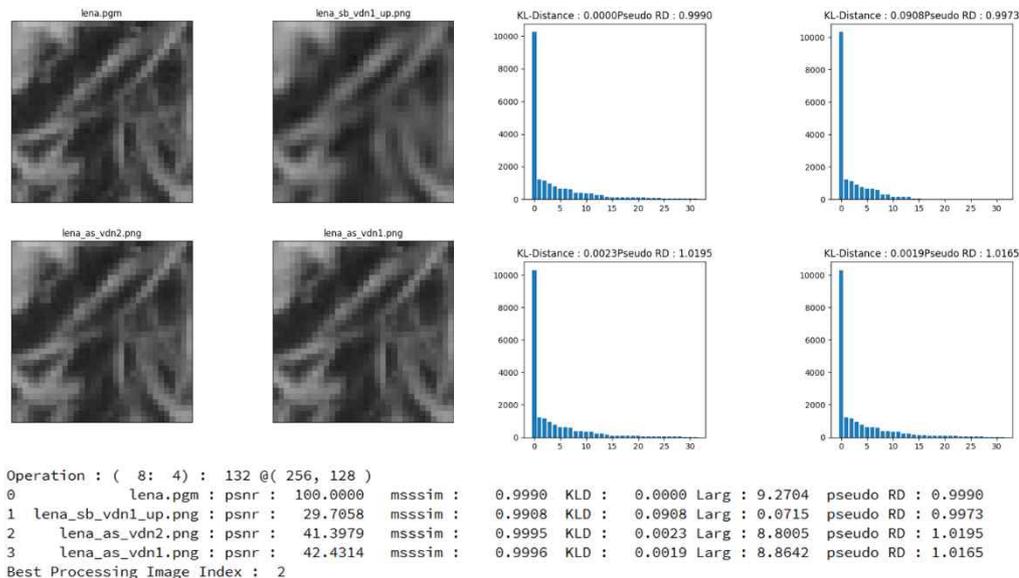


그림 4. 영상처리 필터 2번(as\_vdn2)이 선택되는 경우의 제안한 의사 율-왜곡 함수의 특성  
 Fig. 4. The characteristics of the proposed pseudo rate-distortion function when the image processing filter No.2 (as\_vdn2) is selected

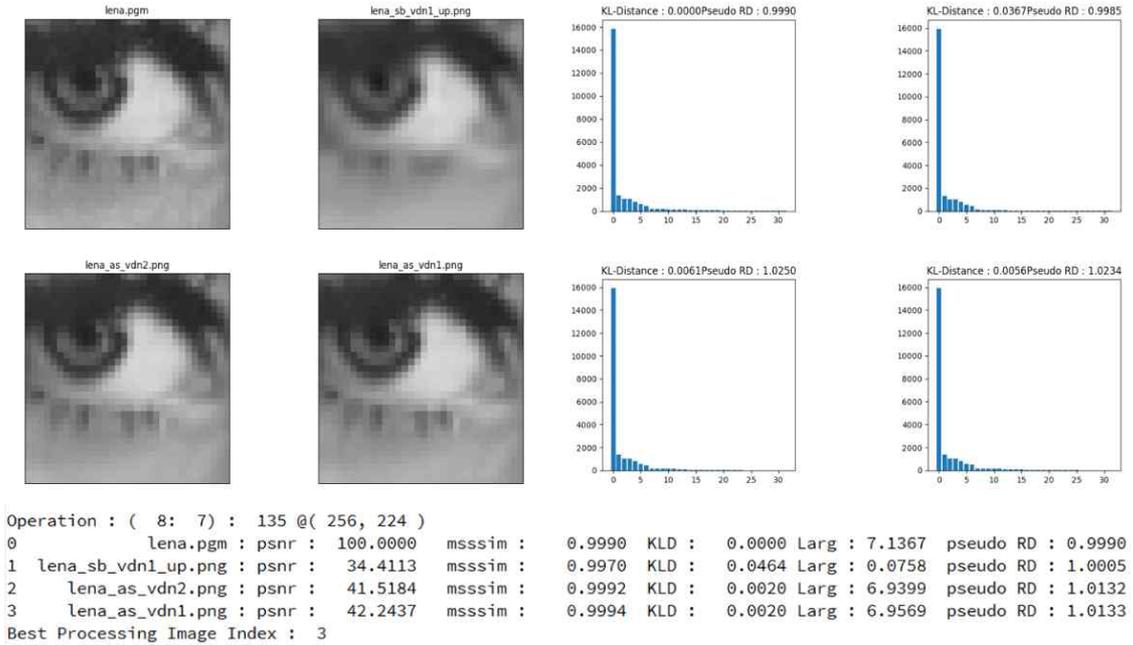


그림 5. 영상처리의 필터 3번(as\_vdn1)이 선택되는 경우의 제안한 의사 율-왜곡 함수의 특성  
 Fig. 5. The characteristics of the proposed pseudo rate-distortion function when the image processing filter No.3 (as\_vdn1) is selected

$m$ 보다 작게 되면 라그랑지안 승수가 작아져 MSSSIM을 최대로 할 수 있는 필터를 선택할 확률이 높아지며, 반대의 경우에는 비트량을 줄일 수 있는 필터를 선택할 확률이 높아진다. 이상을 통해 유도된 추정 비트 율 함수를 적용한 율-왜곡 (Rate-Distortion) 측도는 다음과 같이 정의된다.

$$J_{i,t}(f, \lambda, D_{KL}) = f(R_t \bar{P}_{i,t}) + \lambda(E_k) \cdot D_{KL}(p|q) \quad (7)$$

식 (7)에서  $i$ 는 영상처리 필터에 대한 지시 숫자(Index)를 의미하며  $t$ 는 프레임 내  $t$ 번째 블록 지시 숫자를,  $f(X, Y)$ 는 영상,  $X, Y$ 에 대한 블록 별 MSSSIM을 의미한다.

#### IV. 실험결과

제안한 방법의 유효성을 검증하기 위하여 서론에서 논한 바와 같이, 표 1 과 같은 특성을 가진 3가지 영상처리 필터를 사용한다. 32x32 해상도의 블록에 대하여 3가지 영상 필터를 통해 생성된 영상 중 원본 영상과의 비교를, 식 (7)로 제시된 MSSSIM 및 추정 비트 율 함수로 이루어진 율-왜곡 측도를 통해 가장 큰 값을 가진 처리 영상을 선택하는 방식으로 처리 영상을 만들고 이를 HEVC 참조 소프트웨어 HM 15.1 기반 HEVC (High Efficiency Video Coding) 부호화기의 입력으로 하여 양자화 계수 0에Name of test

표 1. 실험에 사용한 영상처리 필터의 종류와 특성  
 Table 1. Varieties and characteristics of image processing filters in the experiments

Filter 이름	Filter 특성	효과
sb_vdn1_up	Noise-canceling based on wavelet transform after downsampling	Large bit reduction, large PSNR decreasing. No RD gain. Strong artifact canceling
as_vdn2	Noise-canceling based on strong wavelet transform	Appropriate strong artifact canceling, bit reduction of about 20~30% approximately (In the viewpoint of MSSSIM)
as_vdn1	Noise-canceling based on weak wavelet transform	Appropriate artifact canceling, bit reduction of about 20% approximately (In the viewpoint of MSSSIM)
None	original Image	As an original image itself

서 37까지 부호화한 결과를 비교한다. 영상 필터에서 사용된 웨이블릿은 Cohen-Daubechies-Feauveau 9/7 필터에서 사용되는 것으로 이는 JPEG2000의 손실 압축에서 사용되는 웨이블릿이다. 본 논문에서 제안한 알고리즘은 파이썬(Python) 3.5로 구현하였으며 제안한 알고리즘으로 생성된 YUV420P 8bit영상을 HEVC 부호화기 입력으로 사용하였다. 제안하는 알고리즘의 성능을 평가하기 위하여 HEVC 표준화 과정에서 사용되는 공통 시험 조건 (Common Test condition: CTC) RA (Random Access)<sup>[5]</sup> 실험 조건에서의 압축 성능과 비교하였으며, 공통 시험 조건에서 사용되는 테스트 영상 중 Class B에 해당되는 영상들을 대상으로 성능 측정이 수행되었다. Class B 영상의 상세 정보는 표 2와 같다.

표 2. 제안한 알고리즘의 성능 검증을 위해 사용된 HEVC 표준 테스트 영상의 상세 정보

Table 2. Detailed information of the standard HEVC test videos used to verify the performance of the proposed algorithm

Class (Resolution)	Name of test sequence		Number of frames	Frame rate
	Code	Name		
Class-B (1920x1080)	T01	Kimono	32	24
	T02	Park run	32	24
	T03	Cactus	32	50
	T04	Basketball	32	50
	T05	Terrace	32	60

표 4는 Class-B 영상에 대한 양자화 계수 0에서 37까지의 실험 결과이다. 실험결과, 낮은 양자화 계수에서는 제안한

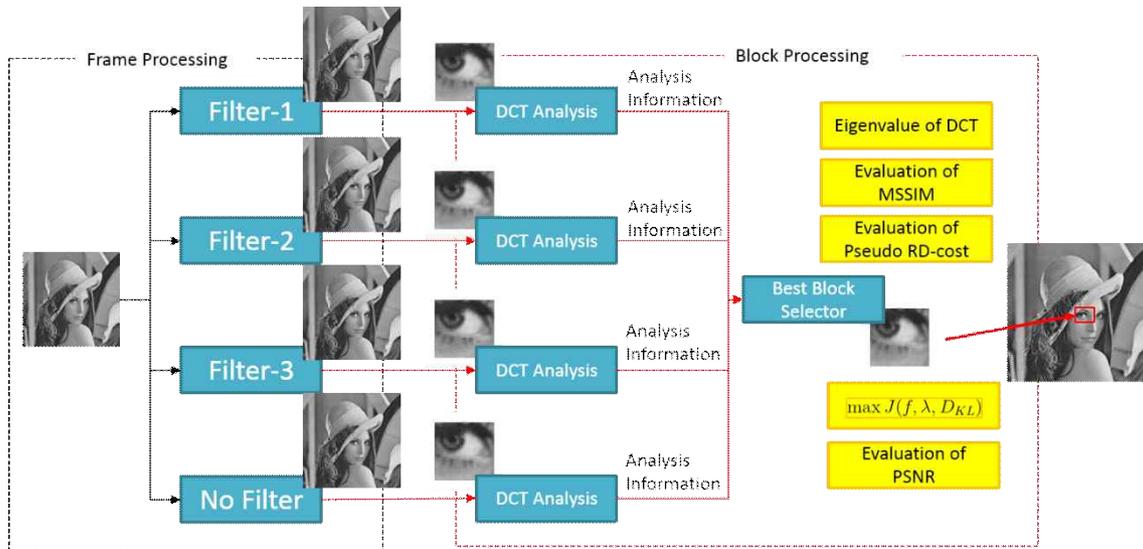


그림 6. 실험에 사용된 제안한 영상처리 알고리즘의 구조  
 Fig. 6. The structure of the proposed algorithm employed in the experiments

표 3. 실험에 사용한 영상처리 필터의 종류와 특성

Table 3. Varieties and characteristics of image processing filters in the experiments

		sb_vdn1_up		as_vdn2		as_vdn1		Proposed (Limit PSNR 32)	
		Average. ΔBitrate	Average. ΔMSSSIM	Average. ΔBitrate	Average. ΔMSSSIM	Average. ΔBitrate	Average. ΔMSSSIM	Average. ΔBitrate	Average. ΔMSSSIM
T01	Kimono	-26.0849	-0.0628	-23.6323	-0.06	-15.81	-0.02	-24.32	-0.06
T02	Park run	-37.8334	-0.4438	-17.7785	-0.09	-13.51	-0.05	-28.48	-0.30
T03	Cactus	-49.0845	-0.4682	-22.7518	-0.09	-22.36	-0.09	-36.11	-0.25
T04	Basketball	-47.2371	-0.2695	-29.7547	-0.09	-24.03	-0.06	-39.03	-0.15
T05	Terrace	-54.3278	-0.9310	-13.7998	-0.06	-15.28	-0.07	-18.16	-0.13
Average		-42.9135	-0.4350	-21.5434	-0.0805	-18.1961	-0.0594	-29.2184	-0.1774

표 4. 실험결과

Table 4. Result of experiments

			Original		Proposed (Limit PSNR 32)		Unit (%)			
			Bitrate	MSSSIM	Bitrate	MSSSIM	delta BitRate	delta MSSSIM	Average. ΔBitrate	Average. ΔMSSSIM
T01	Kimono	0	526720.37	0.9999	253490.13	0.9987	-51.8739	-0.1200	-45.66	-0.09
		6	315695.40	0.9998	125838.20	0.9986	-60.1394	-0.1200		
		12	114965.93	0.9991	55443.15	0.9983	-51.7743	-0.0801		
		18	29522.91	0.9981	23962.20	0.9977	-18.8352	-0.0401		
		22	15123.29	0.9971	14243.66	0.9967	-5.8164	-0.0401	-2.98	-0.03
		27	7316.04	0.9939	7087.71	0.9936	-3.1210	-0.0302		
		32	3561.92	0.9867	3504.05	0.9865	-1.6247	-0.0203		
		37	1808.25	0.9742	1783.83	0.9739	-1.3505	-0.0308		
T02	Park run	0	566326.86	1.0000	299227.82	0.9961	-47.1634	-0.3900	-43.54	-0.35
		6	354239.42	0.9998	165205.05	0.9960	-53.3634	-0.3801		
		12	146979.33	0.9992	78520.07	0.9957	-46.5775	-0.3503		
		18	41298.12	0.9978	30123.50	0.9949	-27.0584	-0.2906		
		22	18669.99	0.9965	15277.41	0.9938	-18.1713	-0.2709	-13.42	-0.25
		27	7859.75	0.9931	6615.83	0.9905	-15.8265	-0.2618		
		32	3509.03	0.9862	3085.80	0.9837	-12.0610	-0.2535		
		37	1602.41	0.9733	1480.41	0.9713	-7.6132	-0.2055		
T03	Cactus	0	362919.78	1.0000	205228.01	0.9964	-43.4509	-0.3600	-54.45	-0.34
		6	264455.03	0.9999	120629.27	0.9963	-54.3857	-0.3600		
		12	159239.00	0.9996	63819.99	0.9961	-59.9219	-0.3501		
		18	64525.56	0.9983	25771.01	0.9955	-60.0608	-0.2805		
		22	19922.04	0.9966	11497.71	0.9947	-42.2865	-0.1906	-17.76	-0.17
		27	5576.81	0.9945	4697.73	0.9928	-15.7631	-0.1709		
		32	2622.99	0.9903	2406.24	0.9886	-8.2635	-0.1717		
		37	1366.19	0.9816	1301.82	0.9802	-4.7113	-0.1426		
T04	Basketball	0	320303.45	1.0000	147932.00	0.9975	-53.8150	-0.2500	-64.36	-0.22
		6	220016.19	0.9999	74851.56	0.9974	-65.9791	-0.2500		
		12	120900.60	0.9995	32713.91	0.9971	-72.9415	-0.2401		
		18	35552.22	0.9980	12548.25	0.9966	-64.7047	-0.1403		
		22	10816.71	0.9971	6857.82	0.9961	-36.5998	-0.1003	-13.70	-0.07
		27	3774.23	0.9956	3359.21	0.9948	-10.9962	-0.0804		
		32	1901.22	0.9928	1813.04	0.9922	-4.6383	-0.0604		
		37	1067.45	0.9873	1040.06	0.9869	-2.5659	-0.0405		
T05	Terrace	0	291205.65	1.0000	232209.03	0.9982	-20.2594	-0.1800	-24.78	-0.18
		6	206376.18	0.9999	156120.00	0.9981	-24.3517	-0.1800		
		12	129419.10	0.9997	94900.55	0.9979	-26.6719	-0.1801		
		18	64367.61	0.9990	46442.84	0.9974	-27.8475	-0.1602		
		22	32564.64	0.9977	24154.40	0.9966	-25.8263	-0.1103	-11.54	-0.08
		27	8670.99	0.9954	7562.04	0.9947	-12.7892	-0.0703		
		32	3149.64	0.9922	2978.91	0.9915	-5.4206	-0.0706		
		37	1409.93	0.9858	1380.09	0.9851	-2.1161	-0.0710		
								-18.16	-0.13	

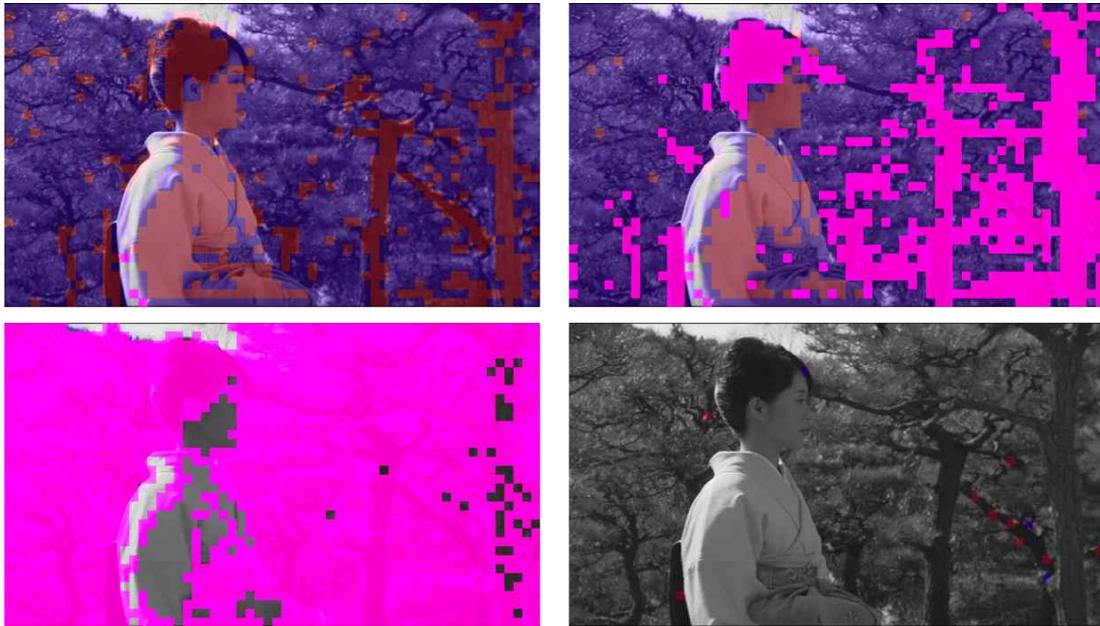


그림 7. 기준 최대 신호 대 잡음비의 변화에 따른 선택 필터의 변화, 남색은 sb\_vdn1\_up, 브라운색은 as\_vdn2, 핑크색은 as\_vdn1, 무색은 원본 영상 의미. 좌상의 경우 기준 최대 신호 대 잡음비 40, 우상은 44, 좌하는 48, 우하는 52.

Fig. 7. The validity characteristics of selecting filters. Indigo blue represents that the filter sb\_vdn1\_up is applied to the block, brown color means the filter as\_vdn2, pink means the filter as\_vdn1, and none color means that the block is the part of original image. For the standard PSNR, it is 40 at the left upper, 44 at the right upper, 48 at the left below, and 52 at the right below

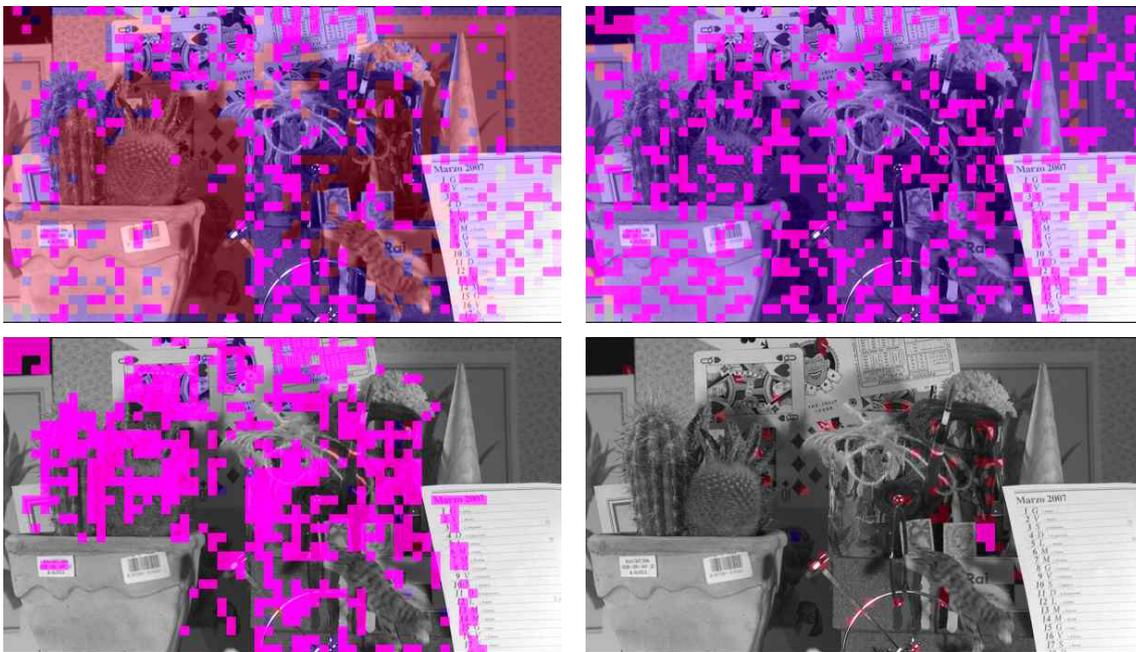


그림 8. Cactus video에 적용했을 때의 경우. 기준 최대 신호 대 잡음비는 좌상 36, 우상 40, 좌하 44, 우하 48

Fig. 8. The same property in the Cactus video set. The standard PSNR is 36 at the left upper, 40 at the right upper, 44 at the left below, and 48 at the right below

알고리즘의 비트 율 절감 효과가 매우 크게 나타나 평균적으로 24~64%의 비트 율 절감 효과가 나타났으며 이때 MSSSIM의 하락율은 0.09~0.35%로 극히 작은 것으로 나타났다, 이에 대한 최대 신호 대 잡음비의 하락은 공통 시험 조건에서 0.28%의 하락을 보였으며 상대적으로 낮은 양자화 계수에서 하락 율이 컸다 한편 공통 시험 조건하의 높은 양자화 계수에서는 영상에 따라 비트 율 절감효과가 다르게 나타나 대체로 2.98~17.76%의 비트 율 절감효과 대비 MSSSIM 하락율은 0.03~0.25%로 나타남을 알 수 있다.

그림 7, 8은 기준 최대 신호 대 잡음비의 변화에 따른 영상 필터 선택의 변화를 보인다. 기준 최대 신호 대 잡음비가 낮은 경우는 비트 율을 저감할 수 있는 영상 필터가 주로 선택되고 기준 최대 신호 대 잡음비가 높아지면 원본영상과의 차이를 줄일 수 있는 영상 필터가 주로 선택된다 기준 최대 신호 대 잡음비는 의사 율-왜곡 함수에 기반한 복수 영상 필터를 선택의 결과 최소 기준 최대 신호 대 잡음비보다 높아지도록 한다.

## V. 결 론

본 논문에서는 서로 다른 영상처리 필터를 통해 생성되는 영상을 조합하여 주관적 화질 하락을 최소화 하면서 비트 율 절감을 이루기 위한 방법론을 제시하였다. 서로 다른 영상처리 필터를 선택하기 위하여 MSSSIM 과 쿨백 라이블러 발산기반의 율-왜곡 측도 함수를 제안하였으며 이를 통해 선택된 영상처리 함수에 의해 생성된 영상은 비트 율 절감과 주관적 화질 하락 최소화라는 두 가지 목표를 달성할 수 있음을 실험적으로 보였다.

본 논문에서 제안하는 방법은 영상 전 처리에 적용될 수 있는 분야이므로 직접적으로 영상 압축에 바로 적용할 수는 없다. 그러나, 본 논문에서 제시한 방법론을 기반으로

영상의 전 처리 대신 예측 영상 중 특히 부 화소 생성과 관련된 부분이나 최소 가지 차이 방법을 사용하여 이산 역 변환 계수를 적절히 제거하는 부분에서 본 논문이 제시하는 방법을 사용할 경우 효과적인 압축 성능 향상을 제시할 수 있을 것으로 판단된다. 특히 본 논문에서 제시하는 율-왜곡 측도 함수는 연속 함수의 특성을 가지고 있기 때문에 향후 기계학습 알고리즘에 있어 동영상 압축 알고리즘 내에 영상 처리 혹은 생성을 위한 기계학습 방법론을 유도하는데 기여할 수 있으리라 생각된다.

## 참 고 문 헌 (References)

- [1] T. Dumas, A. Roumy, C. Guillemot "Image compression with stochastic winner-take-all auto-encoder". Proceeding of International Conference Acoustic, Speech and Signal Processing, New Orleans, USA, pp. 1512-1516
- [2] K. Gregor, Y. LeCun "Learning representations by maximizing compression". arXiv:1108.1169, Aug. 2011
- [3] C. Chou, 1995, A Perceptually Tuned Subband Image Coder Based on the Measure of Just-Noticeable-Distortion Profile, IEEE Transactions on Circuits and Systems for Video Tech, vol. 5, Issue 6, pp.467-476, Dec. 1995.
- [4] Wang, Z. Simoncelli, E.P. Bovik, A.C. "Multiscale structural similarity for image quality assessment". Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, Vol. 2, pp. 1398 - 1402, Feb. 2004. doi:10.1109/ACSSC.2003.1292216.
- [5] F. Bossen, "Common test conditions and software reference configurations," The 8th JCT-VC meeting, JCT-VC H1100, San Jose, CA, Jan. 2012.
- [6] Y. Dai, D. Liu, and F. Wu, "A Convolutional Neural Network Approach for Post-Processing in HEVC Intra Coding," Proceeding of the 23rd International Conference on Multimedia Modeling, Reykjavik, Iceland, pp.28-39, Jan. 2017.16
- [7] T. Wang, M. Chen, and H. Chao, "A Novel Deep Learning-Based Method of Improving Coding Efficiency from the Decoder-end for HEVC," Proceeding of Data Compression Conference, Snowbird, USA pp.410-419, April 2017.
- [8] Brian F. Doolin, Clyde F. Martin, Introduction to Differential Geometry for Engineers, New york, pp. 115-139, 1990.

저 자 소 개



석진욱

- 1995년 2월 : 홍익대학교 전자공학과 공학석사
- 1998년 8월 : 홍익대학교 전자공학과 공학박사
- 2000년 ~ 현재 : 한국전자통신연구원 실감AV연구그룹 책임연구원책임연구원
- 2006년 ~ 현재 : 과학기술연합대학원대학교 겸임교수
- ORCID : <https://orcid.org/0000-0001-5318-1237>
- 주관심분야 : 영상처리/압축, 비선형 확률제어, 기계 학습



조승현

- 2003년 8월 : 경북대학교 전자전기공학부 학사
- 2006년 2월 : 한국과학기술원 전기및전자공학과 석사
- 2015년 8월 : 한국과학기술원 전기및전자공학과 박사
- 2006년 6월 ~ 현재 : 한국전자통신연구원 실감AV연구그룹 책임연구원
- ORCID : <https://orcid.org/0000-0003-1985-4420>
- 주관심분야 : 영상처리/압축, 기계 학습, 디지털 회로설계



김휘용

- 2004년 2월 : 한국과학기술원 전기및전자공학과 박사
- 2003년 8월 ~ 2005년 10월 : (주)에드팩테크놀로지 기술연구소 멀티미디어 팀장
- 2006년 9월 ~ 2010년 8월 : 과학기술연합대학원대학교 겸임교수
- 2013년 9월 ~ 2014년 8월 : Univ. of Southern California 방문연구원
- 2005년 11월 ~ 현재 : 한국전자통신연구원 미디어연구본부 실감AV연구그룹장
- ORCID : <https://orcid.org/0000-0001-7308-133X>
- 주관심분야 : 영상처리/압축, 컴퓨터 비전, UHD/HDR/3D/VR, MPEG



최진수

- 1990년 2월 : 경북대학교 전자공학과 공학사
- 1992년 2월 : 경북대학교 전자공학과 공학석사
- 1996년 2월 : 경북대학교 전자공학과 공학박사
- 1996년 5월 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : <https://orcid.org/0000-0003-4297-5327>
- 주관심분야 : 영상처리/압축, UHDTV 방송, 3DTV 방송