

## HEVC 부호화 부가정보를 이용한 장면전환 검출 연구

엄유미, 박상일, 유성근

서울과학기술대학교

yumieeom@seoultech.ac.kr, {sangilparkmail, orcogre}@gmail.com,

## An analysis of Scene Change Detection using HEVC coding additional information

Yumi Eom, Sangil Park and Sunggeun Yoo  
Seoul National University of Science and Technology

## 요 약

대용량 콘텐츠 수요와 공급의 증가에 따라 UHD 비디오의 분석, 색인, 편집 등을 위한 장면전환을 검출하는 방법이 요구되고있다. 따라서, 이전까지 많은 연구자들이 다양한 장면전환 검출 방법을 연구해왔다. 그러나 카메라의 다양한 움직임과 장면의 변화를 정확하게 검출하기는 어렵다. 또한, 4K (3820x2160) 해상도 이상의 UHD 비디오들은 데이터 량을 더욱더 증가시키기 때문에 이전의 장면전환 검출 방법은 UHD 비디오 콘텐츠에 적용하기에는 너무 많은 시간이 걸리는 문제점이 발생한다. 본 논문에서는 차세대 고효율 코덱 HEVC의 부호화 부가정보를 이용한 4가지 장면전환 검출 방법을 제안한다. 연산량을 줄이기 위해 프레임 특징에 따른 방법을 각각 제시하며, 실험 결과를 통해 HEVC로 부호화 된 UHD 콘텐츠들의 장면전환 검출 가능성을 확인한다.

## 1. 서론

2013 년 11 월 ISO/IEC 는 차세대 코덱 HEVC 표준을 발표했다 [1]. 또한 LG 는 7680x4320 해상도를 가지는 8k UHD TV 를 발표했고 삼성은 갤럭시 노트 4 는 HEVC 디코더를 탑재하여 출시했다. Hevc 는 H.264/AVC 에 비해 3 배 이상 복잡하지만 2 배이상의 압축률을 제공한다 [2]. 따라서, UHD 비디오 콘텐츠가 나날이 늘어가는 요즘, 대용량 데이터의 저장, 관리, 편집 등을 편리하게 하기 위해서는 압축률이 높은 HEVC 가 필수적으로 사용된다. 또한 비디오 데이터의 관리에는 색인, 검색, 편집을 가능하게 하는 장면전환 정보가 필수적이라고 말할 수 있다. Davinci Resolve 와 같은 편집 소프트웨어는 H.264/AVC 를 지원하여 장면전환을 검출하여 편집에 도움을 주는 기능이 내장되어 있다. 하지만 이 기능은 전체 복호화를 해야하므로 대용량 비디오 편집 시에는 이 기능이 매우 느리기 때문에 실제로 사용하기 어렵다. 게다가 HEVC 는 H.264/AVC 에 비해 3 배 이상 복잡하므로 이는 전체 복호화로 실시간 장면전환 검출하는 것이 불가능할 것이다. 따라서, UHD 콘텐츠를 전체 복호화하지 않고 장면전환을 검출하는 방법이 필수적이라고 볼 수 있다.

본 논문에서는 HEVC 복호화시에 생성되는 부가정보를 이용하여 장면전환을 검출하는 방법을 제안한다. 구체적으로 프레임의 특징에 따라 I-프레임과 P/B 프레임으로 나누어 장면전환을 검출하는 방법을 제시하고, HEVC 비트스트림 상에서 장면전환 검출의 가능성을 살펴본다.

본 논문의 구성은 다음과 같다. 2 절에서는 HEVC 부호화 부가정보를 장면전환 검출과 관련하여 살펴본 후, 3 절에서는 I-프레임 부가정보를 장면전환 검출에 적용하는 방법을

설명하고, 4 절에서는 P/B-프레임 부가정보를 장면전환 검출에 적용하는 방법을 설명한다. 5 절에서는 제안한 방법의 실험결과를 확인한다. 마지막으로 6 절에서는 본 논문에 대한 결론을 맺는다.

## 2. HEVC 부호화 부가정보와 장면전환 검출

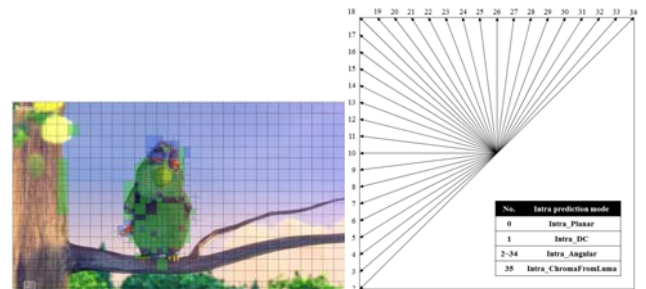


그림 1. (좌) Coding unit structure of HEVC

그림 2. (우) Intra prediction mode of HEVC

HEVC 에서 무엇보다 가장 큰 특징은 가변크기 부호화 유닛을 사용하여 압축효율을 좋게 한다는 점이다. 부드러운 영역은 사이즈가 큰 블록, 디테일한 영역에는 사이즈가 작은 블록을 사용함으로써 압축 효율을 높인다. 이러한 가변크기 부호화 유닛을 사용하여 그림 1 과같이 프레임 내의 구조적 특징을 자세히 나타내어 줄 수 있다. 프레임 내의 구조적 특징은 장면전환 검출에 있어 중요한 요소가 될 수 있다. 또한 장면전환 상황이 일어나면 대부분의 CU(Coding Unit)는

인트라 모드로 부호화되므로, CU 의 크기, 위치 뿐 아니라 예측모드의 비율도 장면전환 검출에 중요한 요소가 될 수 있다.

또 하나의 HEVC 특징중 하나는 그림 2 에서 보이는 바와 같이 인트라 예측 모드의 개수가 35 가지로 늘어났다는 점이다. 장면전환 상황이 일어나면 프레임 내의 특징이 달라지므로 PU(Prediction Unit)의 인트라 예측 모드 분포가 달라지는 것을 예상할 수 있다. 또한, 벡터예측도 AMVP(Adaptive Motion Vector Prediction), 움직임 병합 모드로 발전했으므로 이러한 특징들을 이용한다면, 이전 장면전환 검출 방법을 응용하는 것 뿐만 아니라, 새로운 HEVC 부호화 기반의 장면전환 검출기법을 개발하는 것도 가능해 보인다.

그러나, 대용량 비디오 콘텐츠의 경우 압축효율이 좋은 HEVC 를 사용해도 연산복잡도가 매우 높기 때문에 모든 프레임을 복호화하여 장면전환 검출을 시행하기에는 한없이 많은 연산량을 가지고 있다. 따라서, 최소한의 복호화에 따른 부가정보를 이용하여 장면전환 검출을 하기 위해서는 프레임 특징에 따른 장면전환 검출 방법이 필요하다.

일반적으로 프레임은 I/P/B-프레임으로 나누어질 수 있는데 I-프레임은 다른 프레임들과는 독립적으로 부호화 된다. I-프레임은 시간적으로 앞, 뒤의 프레임을 참조하지 않으므로 움직임 예측을 적용하지 않으며 화면 내 정보만을 사용해서 부호화하므로 프레임 내의 픽셀간의 상관관계가 중요한 정보로 이용된다. P-프레임은 이전의 I-프레임과 P-프레임을 참조하여 부호화 된다. I 프레임과 다르게 참조 프레임이 있으므로 정지된 부분을 제외한 움직임이 있는 부분만 부호화되며, 현재 프레임과 과거 프레임간의 상관관계가 중요한 정보로 이용된다. 반면에 B-프레임은 이전의 I/P-프레임 뿐만 아니라 이후의 I/P-프레임도 참조하여 부호화 한다. B 프레임의 경우 참조하는 프레임이 많아, 중복요소를 가장 많이 제거할 수 있으므로 압축률이 가장 높다 [3]. 따라서, P/B-프레임은 화면간 부호화, 화면 내 부호화, 스킵 모드 총 세 가지에 대한 부호화 과정을 거치고, I-프레임은 화면 내 예측모드에 대한 부호화 과정만 거치게 된다. 또한, 복호화시 복잡도는 인코딩의 복잡도와 상관관계가 높는데, I-프레임의 경우 복호화 복잡도가 P/B-프레임보다 상대적으로 낮으므로 I-프레임은 전체 복호화, P/B-프레임은 최소한으로 복호화하여 장면전환 정보를 찾는다면 빠른 검출이 가능할 것으로 보인다.

### 3. HEVC 부호화 부가정보를 이용한 I-프레임 장면전환 검출

#### A. Intra Prediction Mode and CU Size:

HEVC 에서는 가변적으로 크기조절이 가능한 부호화 유닛 CU 를 제공하므로 다양한 CU 의 구성 정보는 장면전환 검출의 정보로 이용될 수 있다. 장면전환이 일어나지 않는다면 CU 의 구성정보는 완전히 같을 수는 없지만 그 유사도는 높을 것이다. 또한 35 가지 인트라 예측 모드도 장면전환 되지 않은 전, 후 프레임에서는 유사할 것으로 예상된다. 그러므로, CU 크기에 따른 인트라 예측 모드의 개수의 차를 이용한다면 장면전환 검출이 가능할 것으로 보인다 [4].

$$\frac{\sum_{i=0}^C \sum_{j=0}^M (P(n)_{ij} - P(n-1)_{ij})^2}{cnt(CU_n) + cnt(CU_{n-1})} > T \quad (1)$$

식 (1)에서 현재 I-프레임을 n 번째 프레임이라고 하고 바로 전 I-프레임을 n-1 번째 프레임이라고 한다면, C 는 5 가지 CU 의 크기, M 은 35 가지의 인트라 예측 모드 정보를 의미한다. 제안하는 방법에서 CU 의 크기는 4x4, 8x8, 16x16, 32x32, 64x64 인 5 가지를 이용하였다. Cnt(CU<sub>n</sub>)는 n 번째 프레임 안의 총 CU 수, Cnt(CU<sub>n-1</sub>)는 n-1 번째 프레임 안의 총 CU 수를 의미한다. CU 크기에 따른 인트라 예측 모드 개수의 SSE(Sum of Squared Error)를 인접한 두 프레임 CU 개수의 총 합으로 나눈 값이 임계값 T 보다 크면 장면전환이라고 판단한다. 표 1 에서는 이 실험의 결과값을 보여주고 있다.

표 1. INTRA PREDICTION MODE AND CU SIZE

No.	Total Frame	Total SCD	T	N <sub>C</sub>	N <sub>F</sub>	N <sub>M</sub>
A	208	28	10	22	10	1

#### B. Intensity and Sums of Squared Error:

장면전환이 되지 않은 프레임 간에서는 프레임이 가지는 특징이 비슷하기 때문에 인텐시티 값도 비슷한 양상을 가질 가능성이 높다. 장면전환이 일어날 경우 프레임 간의 픽셀단위 인텐시티의 차이값이 확연히 드러날 것으로 보인다. 제안하는 방법은 대용량 비디오의 데이터를 1/16 로 공간적 서브 샘플링을 하여 장면전환 검출이 가지는 복잡도를 낮췄다.

$$If \ D = \sqrt{(P(n)_{ij} - P(n-1)_{ij})^2} > 35 \quad (2)$$

Then: 1, Else: 0

$$\sum_{i=0}^H \sum_{j=0}^W D_{ij} > T \quad (3)$$

식 (2)에서 현재 I-프레임을 n 번째 프레임이라고 하고 바로 전 I-프레임을 n-1 번째 프레임이라고 한다면, D 는 인접한 두 프레임의 동일한 위치에 있는 픽셀의 인텐시티 차이 값이다. 연산 복잡도를 낮추기 위해 차이값 D 가 35 보다 크면 1, 작으면 0 으로 이진화를 하기로 한다. 이진화 된 인접한 두 프레임간의 픽셀단위 인텐시티를 모두 더한 값이 임계값 T 보다 크다면 장면전환으로 판단한다. 표 2 와 그림 3 에서는 이 실험의 결과값을 보여주고 있다.

표 2. INTENSITY AND SUMS OF SQUARED ERROR:

No.	Total Frame	Total SCD	T	N <sub>C</sub>	N <sub>F</sub>	N <sub>M</sub>
B	208	28	1,000,000	26	0	2

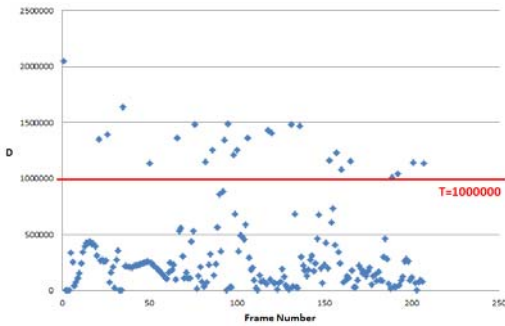


그림 3. A distribution chart of sum of squared error

#### 4. HEVC 부호화 부가정보를 이용한 P/B-프레임 장면전환 검출

##### A. Intra, Inter, Skip mode and Threshold:

HEVC 에서 단위 블록들은 인트라, 인터, 스킵모드로 나뉘어 부호화 될 수 있다. 인트라 모드는 화면 내 중복성을 제거하고 인터 모드는 화면 간 중복성 제거, 스킵모드는 이전 프레임과 비슷하므로 보조정보를 부호화 하지 않는다. 인트라, 인터, 스킵모드의 특징을 이용하면 장면전환이 일어나는 프레임의 경우 인트라모드 블록이 눈에 띄게 많고, 인터, 스킵모드 블록이 현저하게 낮다는 것을 예상할 수 있다 [5].

$$S_n = \frac{Intra+1}{Inter+Skip+1} > T \quad (4)$$

식 (4)에서  $S_n$  은 n 번째 프레임의 인트라, 인터, 스킵모드의 비율이다. 1 은 영이나 무한대 값이 나오지 않기 위한 상수이다.  $S_n$  이 임계값 T 보다 크면 장면전환으로 판단한다. 표 3, 그림 4 에서는 이 실험의 결과값을 보여주고 있다.

표 3. INTRA, INTER, SKIP MODE AND THRESHOLD

No.	Total Frame	Total SCD	T	N <sub>C</sub>	N <sub>F</sub>	N <sub>M</sub>
C	9768	24	180	23	0	1

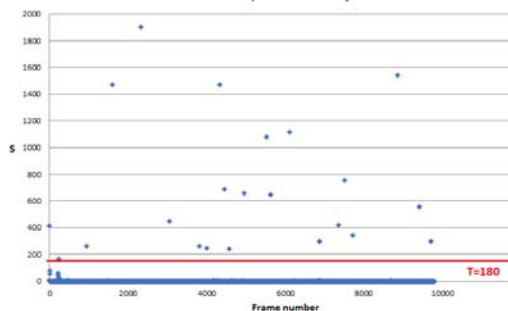


그림 4. The ratio of Intra, Inter and Skip mode

##### B. Intra, Inter, Skip mode and Window Size:

장면전환을 판단하기 위해 임계값을 사용한다면 임계값을 정하는 방법이 추가로 필요하다. 임계값을 정하는 방법은 반복 실험을 통해서 값을 얻는 실험적 방법, 동영상의 전체 프레임을 분석해서 값을 얻는 분석적 방법, 동영상의 정보를 가지고 가변적으로 값을 얻는 적응적 방법 세가지가 있다. 우리는 자동적 알고리즘에 적용이 가능한 윈도우를 통해 임계값을 얻고자 한다. Feng Jie 등의 논문에 따르면 장면전환이 일어난 프레임의 경우, 그 프레임의 인트라 모드 블록개수가 앞, 뒤 프레임의 인트라 모드 블록 개수를 모두 더한 것 보다도 눈에 띄게 많다 [5].

$$S_n = \frac{Intra+1}{Inter+Skip+1} \quad (5)$$

$$S_w = \sum_{n=n-w}^{n-1} S_n + \sum_{n=n+1}^{n+w} S_n \quad (6)$$

$$S_w > S_n \quad (7)$$

식 (5)에서  $S_n$  은 n 번째 프레임의 인트라, 인터, 스킵 모드의 비율이고, W 는 윈도우 사이즈를 의미한다.  $S_w$  는 n 번째 프레임을 제외하고 윈도우 사이즈 만큼 앞, 뒤 프레임의  $S_n$  값의 합이다.  $S_n$  이  $S_w$  보다 크면 장면전환으로 판정한다. 표 4 는 윈도우 사이즈에 따른 장면전환 검출 결과값을 보여준다.

표 4. INTRA, INTER, SKIP MODE AND WINDOW SIZE

No.	Total Frame	Scene Change	W	N <sub>C</sub>	N <sub>F</sub>	N <sub>M</sub>
D	9768	24	20	22	14	2
			25	22	6	2
			30	22	6	2
			35	22	4	2
			40	22	3	2
			45	22	3	2

#### 5. 실험 결과

실험에 이용된 PC 는 Intel I7 4770 에 32GB ram 을 장착하였고, 이미지 처리를 위해 HM16.2(HEVC test Model)와 Opencv 2.4.9 를 연동하였다. 또한, 결과와 비교할 정확한 장면전환 정보를 찾기 위해 Elecard 사의 HEVC analyzer 를 이용하였다. 실험을 위한 샘플 비디오는 <https://media.xiph.org> 에서 제공되는 bbb 이며, 1920x1080p, YUV 4:2:0, 8bits, GOP(Group of picture)는 48 이며 사용된 영상들은 IBBBBBIBBBBB 의 구조를 갖는다.

제안하는 방법들의 성능평가는 장면전환 검출 의 척도로 널리 알려진 precision, recall, F1 을 사용하였다.

$$Precision = \frac{N_C}{N_C + N_F} \quad (8)$$

$$Recall = \frac{N_C}{N_C + N_M} \quad (9)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (10)$$

식(8)–(10)은 정확도를 나타내기 위한 척도로  $N_C$  는 정확하게 검출한 장면전환 개수,  $N_F$  는 잘못 검출한 장면 전환 개수,  $N_M$  은 검출하지 못한 장면 전환 개수를 의미한다. 따라서 precision 은 임의의 알고리즘에서 장면전환 검출을 수행한 결과 중 정확한 검출 비율을 의미하며, recall 은 실험 영상의 실제 장면 전환 결과 중에서 해당 알고리즘이 정확하게 검출한 비율을 의미한다. F1 은 정확도를 판단하는 종합적인 척도로써 precision 과 recall 의 조화평균으로 정의된다. 이를 이용하면 precision 이나 recall 중 하나의 척도만 높고 다른 척도는 낮은 경우의 알고리즘들을 객관적으로 비교할 수 있다.

표 5. RESULT TOTAL

No.	Frame	Scene Change	T/W	Precision	Recall	F1
A	I	28	10	0.69	0.97	0.81
B			1,000,000	1	0.92	0.96
C	P/B	24	180	1	0.96	0.98
D			40	0.88	0.92	0.90

A 는 장면전환 검출 알고리즘으로 사용하기에는 무리가 있다. 그러나, 각각 CU 의 위치정보를 추가적으로 사용하여 검출하면 좋은 결과값을 기대할 수 있을 것 같다. B 의 경우 전체적인 인텐시티가 비슷한 장면전환 부분은 검출하지 못했다. 예를 들어 장면은 전환되었지만 전체적인 인텐시티를 따져보았을 때 비슷하다면 검출하기에 어려움이 따랐다. 따라서 프레임의 컬러 값이나 픽셀의 위치정보를 추가하여 정확도를 높여야 할 것으로 보인다. C 의 경우 가장 높은 검출률을 보였지만 알고리즘을 자동화 하려면, 임계값을 자동적으로 결정하는 알고리즘이 추가적으로 필요하다. D 는 윈도우 기술을 사용하여 자동으로 임계값을 결정했지만 점진적인 장면전환을 찾기 위해서는 누적분포를 이용하는 알고리즘이 추가로 사용되어야 할 것으로 보인다. 4 가지 실험은 공통적으로 점진적인 장면전환을 검출하기 어려웠다. 그러나 각 블록의 위치정보와 누적분포를 추가적으로 이용한다면 좋은 결과값이 기대된다.

## 6. 결론

이 논문은 HEVC 복호화시 부가적으로 생성되는 정보들을

이용하여 장면전환을 검출하는 방법에 대해 제안하였다. 제안하는 방법에서는 프레임의 특징에 따른 부가정보를 이용하였다. 방법 A 는 CU 크기에 따른 인트라 예측모드가 사용되었으며 방법 B 는 프레임의 인텐시티가 이용되었다. 방법 C 와 D 는 인트라, 인터, 스킵 모드 블록들의 비율을 이용하였으며 방법 C 에서는 실험적 임계값을 사용하였고 방법 D 에서는 적응적 임계값인 윈도우를 이용하였다. 4 가지 방법들은 평균적으로 91%의 검출률을 보여주며, HEVC 에서 전체 복호화를 하지 않고 장면전환을 검출할 수 있는 가능성도 함께 보여준다. 결과적으로, 누적분포와 CU 의 위치정보를 추가적으로 적용하는 알고리즘을 이용한다면 HEVC 로 복호화된 대용량 콘텐츠의 장면전환 정보를 빠르고 정확하게 검출할 수 있을 것으로 보인다.

## 참고문헌

- [1] Bross, B., et al. "High efficiency video coding (HEVC) text specification draft 10 (Doc. JCTVC-L1003)." (2013).
- [2] Grois, D., Marpe, D., Mulyoff, A., Itzhaky, B., & Hadar, O. (2013). Performance comparison of H. 265/MPEG-HEVC, VP9, and H. 264/MPEG-AVC encoders. In PCS (pp. 394–397).
- [3] Ewerth, R., & Freisleben, B. (2003, September). Frame difference normalization: an approach to reduce error rates of cut detection algorithms for MPEG videos. In Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on (Vol. 2, pp. II-1009). IEEE.
- [4] Bossen, F., Bross, B., Suhring, K., & Flynn, D. (2012). HEVC complexity and implementation analysis. Circuits and Systems for Video Technology, IEEE Transactions on, 22(12), 1685–1696.
- [5] Jie, F., Aiai, H., & Yaowu, C. (2008, December). A novel scene change detection algorithm for H. 264/AVC bitstreams. In Computational Intelligence and Industrial Application, 2008. PACIA'08. Pacific-Asia Workshop on (Vol. 1, pp. 712–716). IEEE.