

격자 피라미드 GOP 구조 기반의 다시점 비디오 부호화 방법

*오관정, *오한, *호요성, **최병호

*광주과학기술원 정보통신공학과

**전자부품연구원

*{kjoh81, ohhan, hoyo}@gist.ac.kr, **bhchoi@keti.re.kr

Multi-view Video Coding based on Grid-type Pyramid GOP Structure

*Kwan-Jung Oh, *Han Oh, *Yo-Sung Ho, **Byeong-ho Choi

*Gwanju Institute of Science and Technology (GIST)

**Korea Electronics Technology Institute (KETI)

요약

디지털 멀티미디어 시대를 맞이하여 영상통신 기술이 급속히 발전함에 따라 보다 사실감과 몰입감을 줄 수 있는 3차원 입체 영상처리에 대한 관심이 커지고 있다. 최근 국내의 연구기관에서 다차원 멀티미디어 서비스 개발을 위한 연구가 활발히 진행되고 있으며, MPEG 표준화 그룹에서도 H.264/AVC 압축 방법을 이용한 다시점 비디오 부호화(multi-view video coding, MVC) 방법들이 제안되었다. 본 논문에서는 격자 피라미드 GOP 구조 기반의 다시점 비디오 부호화 방법에 대해 기술하였다. 이 방법은 현재 MPEG 표준화 그룹에서 권고된 'Anchor' 방법에서 고려치 못한 인접 시점간의 공간적인 상관도를 효과적으로 활용하기 위해 격자 GOP 구조를 제안했고, 각 시점에 대한 효율적인 부호화를 위해 계층적 피라미드 GOP 구조를 이용하였다. 또한, 공간적인 예측의 경우에 시점간의 전체 변이(global disparity)를 고려하여 가변적인 탐색 범위를 이용하였다. 본 논문에서 제안한 방법은 현재 MPEG에서 성능 평가의 기준이 되는 'Anchor' 방법에 비해 동일 비트율에서 0.5~0.8 dB 정도의 성능 향상을 보였다.

1. 서론

다시점 비디오(multi-view video) 영상이란 똑같은 3차원 장면을 두 대 이상의 카메라를 이용하여 촬영한 영상들로, 기하학적인 교정을 거친 여러 영상들의 공간적인 합성을 통해 사용자에게 다양한 시점의 영상을 제공할 수 있다.

다시점 비디오 영상의 한 예인 파노라마(panoramic) 영상은 우주/항공 사진학, 컴퓨터 비전, 컴퓨터 그래픽스 분야에서 많이 연구되고 있으며, 항공사진의 해석, 영상변화 감지, 비디오 압축, 비디오 인덱싱, 카메라 해상도 및 FOV(field of view) 확대에서 간단한 영상 편집에 이르기까지 매우 다양한 분야에 응용되고 있다. 컴퓨터 비전에서는 다시점에서 획득된 여러 영상을 이용하여 영상내의 물체의 깊이(depth)와 시차(disparity) 정보를 추출하고 있으며, 컴퓨터 그래픽스에서도 영상기반 렌더링(image based rendering)이란 이름으로 획득된 다시점 영상들을 이용하여 가상의 시점에서 사실적인 영상을 생성한다.

이러한 다시점 비디오 처리기술은 전방향성 카메라를 이용한 감시 시스템이나, 게임에서 이용되는 3차원 가상 시점, 또는 다수의 카메라 영상들로부터 입력된 영상을 임의로 선택할 수 있도록 하는 시점 스위칭 등에 이용되고 있다. 또한, 이러한 다시점 비디오 영상은 네트워크 기술과 맞물려 대화형 콘텐츠나 실감 콘텐츠를 이용하는 다양한 멀티미디어 서비스에 확장될 수 있다.

이처럼 다시점 비디오 영상은 사용자에게 자유로운 시점 및 넓은

화면을 통한 입체감 제공이라는 큰 장점을 가지고 있지만, 다수의 카메라를 이용한 영상 획득이 쉽지 않고, 데이터양이 너무 방대하기 때문에 실시간으로 처리할 데이터양의 증가 및 고가의 장비가 요구되는 문제점 때문에 다양한 서비스 개발이 상당히 제한되고 있다 [1].

현재 MPEG-4 3DAV 표준화 작업의 기준이 되는 다시점 영상 부호화 방법은 각 시점의 비디오 영상을 최신 비디오 압축 표준인 H.264를 이용하여 독립적으로 부호화하는 방법을 취하고 있다. 그러나 이 방법은 다시점 비디오의 중요한 특성 중의 하나인 인접 시점 영상들 사이의 유사성인 공간적인 상관도를 전혀 고려치 않았다. 인접한 시점의 경우 똑같은 3차원 장면을 촬영하는 카메라의 위치가 조금 바뀐 것이기 때문에 동일 시간의 인접한 시점의 영상들은 상당히 유사하므로 실제로 이를 잘 이용하면 부호화 효율을 크게 향상시킬 수 있다.

본 논문에서는 다시점 비디오 영상의 방대한 데이터를 효율적으로 압축하는 방법을 제안한다. 이는 다음과 같은 세 가지의 세부 알고리즘으로 나뉜다. 첫째로, 부호화 효율에 큰 영향을 미치는 I 화면의 영향을 다시점 영상 전체에 효율적으로 전달하기 위한 격자 GOP 배치 알고리즘을 제안한다. 둘째로, 각 GOP에 대한 효율적인 부호화를 위해 RB 화면과 B 화면을 이용한 피라미드 GOP 구조를 제안한다. 마지막으로, 효율적으로 공간적인 상관도를 이용하기 위해 인접 시점의 화면을 참조 화면으로 이용하는 경우에 대해서는 화면의 전체 변이를 고려하여 움직임 예측에 이용되는 탐색 범위를 2배로 확장하는 가변적인 탐색범위 알고리즘을 제안한다.

2. 다시점 비디오 부호화

가. 다시점 비디오 부호화의 기준 방법

MPEG-4 3DAV 그룹에서는 현재 작업중인 다시점 비디오 부호화 기술들의 성능 비교를 위해 H.264/AVC를 이용하여 여러 시점의 영상들은 각각 독립적으로 부호화하는 방법을 'Anchor'라고 설정하고 있다. Anchor 부호화 방법은 입력 영상의 초당 프레임수에 상관없이 IBBP 형태의 GOP 구조를 취하고 있으며, 비트-왜곡 최적화와 루프 필터를 이용한다. 또한, 가변길이부호화에는 CABAC (context based binary arithmetic coding)을 이용한다. 움직임 예측의 탐색 범위는 ± 32 이고, 참조 화면은 5개를 이용한다. I 화면의 간격은 0.5초가 기준이다. IDRIntraEnable를 사용하며, 비트율 조절을 위해 비트율 제어를 이용한다. 그림 1은 초당 25프레임의 8시점을 영상에 대한 Anchor 방법의 한 GOP에 대한 부호화 구조를 보여주고 있다.

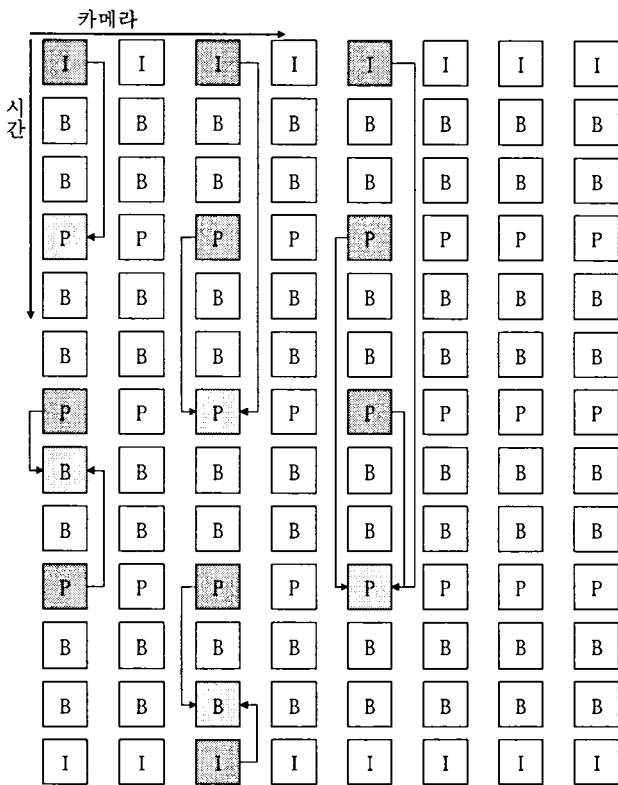


그림 1. 'Anchor' 방법의 부호화 구조도

그림 1에 보인 것처럼, Anchor 방법의 경우 각 시점에 대한 부호화는 독립적으로 이루어지며, GOP 구조는 IBBP 구조를 반복적으로 사용한다. 여기서 화살표는 화면 사이의 참조 관계를 보여주고 있다. 그림 1에서 노란색 화면은 참조 화면을 이용해 부호화된 화면을 의미하고, 녹색 화면은 노란색 화면을 부호화하는데 이용된 참조 화면을 나타낸다. 그림 1에서 보듯이, P 화면은 이전에 부호화 된 I 화면이나 P 화면을 참조 화면으로 이용하여 부호화된다. B 화면은 화면상 앞뒤에 위치한 P 화면이나 I 화면들을 이용하여 부호화된다. 따라서 Anchor 방법에서는 인접 시점에 대한 공간적인 참조는 전혀 이루어지지 않고, 오로지 동일 시점의 시간적인 참조만을 이용한다. I 화면의 위치도 0.5 초 간격인 12 프레임마다 모든 시점에서 동일한 위치에 반복적으로 나오는 규칙적인 GOP 구조를 갖는다.

나. 다시점 비디오 부호화의 요구사항

다시점 비디오 부호화에는 다음과 같은 사항들이 요구된다. 첫째로, 방대한 데이터양을 효과적으로 줄일 수 있는 압축 효율이 요구된다. 둘째로, 시간적 및 공간적인 임의의 접근이 가능해야 한다. 즉, 사용자가 원하는 시점의 원하는 장면에 대한 임의의 접근이 가능해야 한다. 셋째로, 부호화 시간이 짧아야 한다. 순간적으로 처리해야 할 데이터양이 많고 H.264의 자체 복잡성 때문에 부호화 시간이 오래 걸리지만, 가능한 한 짧은 시간에 부호화가 완료되어야 한다. 넷째로, 시점간의 화질이 유사해야 한다. 이밖에도 시점/시간적인 계위성(scalability)이 요구되며, 다양한 해상도와 색차 형태도 지원해야 한다. 추가적으로, 아직 구현이 되지 않았으나, 여러 시점에 대한 병렬처리도 실제 응용에서는 요구되는 사항이다 [2].

3. 새로운 다시점 비디오 부호화

가. 격자 GOP 배치 알고리즘

기존의 Anchor 방법이 이용되는 부호화의 구조를 살펴보면, 모든 시점에 대해 GOP의 배치가 동일한 구조를 갖고 있음을 알 수 있다. 그러나 비디오 부호화의 특성상 I 화면에 가까울수록 좋은 부호화 효율을 보이는 반면에, I 화면에서 멀어질수록 부호화 효율이 떨어진다. Anchor 방법의 경우에는 각 시점에 대한 부호화가 독립적으로 이루어지기 때문에 동일한 GOP의 배치도 무관하다. 하지만, 부호화 효율의 향상을 위해 공간적으로 다른 시점의 영상을 참조할 경우에는, 동일한 GOP 구조보다는 그림 2와 같은 격자 형태의 GOP 배치가 공간적인 상관도를 이용하는데 있어 많은 장점을 갖는다.

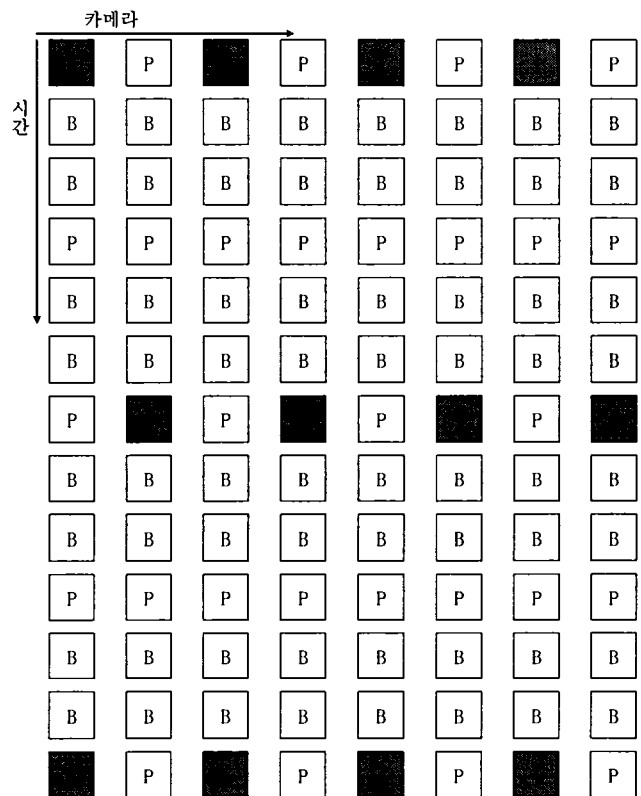


그림 2. 'Anchor' 방법의 부호화 구조도

그림 2는 그림 1과는 달리 I 화면을 격자로 배치하고 있다. 이는 앞에서 언급했듯이 공간적으로 이웃한 시점의 영상을 참조할 경우에 기존의 배치에 비해 고화질의 I 화면의 영향을 다시점 영상 전체에 효율적으로 분배할 수 있다. 즉, 기존의 방법과 동일한 I 화면 간격을 이용하면 부호화 효율을 상당히 향상시킬 수 있고, 기존의 방법에 비해 2배의 I 화면 간격을 이용하면 보다 적은 비트를 이용하여 기존의 I 화면 간격을 이용한 것과 같은 효과를 볼 수 있다. EK라서, 제안한 방법은 기존의 방법에 비해 I 화면 간격을 2배로 늘리는 효과가 있다.

나. 피라미드 GOP 구조 알고리즘

기존의 Anchor 방법에 이용되는 GOP 구조는 IBBP 구조이지만, 여기서 제안한 방법에서는 피라미드 GOP 구조 알고리즘을 이용한다. 피라미드 GOP 구조는 B 화면을 이용한 계층적인 부호화를 통해 한 GOP에 대한 부호화 효율을 극대화시키는 방법이다. 그림 3은 제안한 알고리즘에 이용된 4계층 피라미드 GOP 구조를 보여주고 있다.

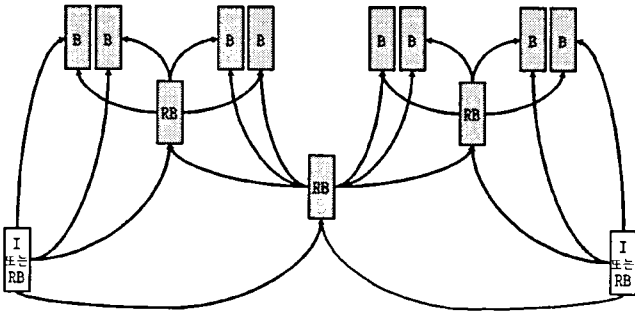


그림 3. 피라미드 GOP 구조

그림 3에 보인 것처럼, 하나의 GOP는 네 계층으로 이루어져 있다. 먼저 최하위 계층은 I 혹은 RB 화면으로 구성된다. 여기서 RB 화면이란 B 화면처럼 부호화되지만, B 화면과 달리 참조 화면으로 사용될 수 있다. 이러한 RB 화면을 이용하여 기존의 P 화면의 자리를 대체했다. 피라미드 GOP의 부호화는 하위 계층부터 계층적으로 이루어진다. 계층 1은 I 화면이나 RB 화면인데, 계층 1의 RB 화면은 동일 시점의 I 화면과 인접 시점의 I 화면을 이용하여 부호화된다. 계층 1의 RB 화면에 대한 부호화시 참조 관계는 그림 5에 나타나 있다. 계층 2의 RB 화면인 계층1의 I 화면이나 RB 화면을 이용하여 부호화된다. 그리고 계층 3의 RB 화면은 계층 2의 RB 화면과 계층 1의 화면을 이용하여 부호화된다. 같은 방법으로, 계층 4의 B 화면들은 이전 계층의 화면들을 이용하여 부호화된다. 피라미드 GOP를 이용하여 부호화하면 기존의 GOP 구조에 비해 보다 적은 비트를 이용한 부호화가 가능하다. 또한 각 계층적인 구조를 이용한 시간적인 계위화도 가능하다 [3].

다. 가변적 탐색 범위 알고리즘

다시점 영상의 부호화 효율 향상을 위해서는 공간적인 상관도를 효율적으로 이용해야 한다. 이를 위해 기존의 많은 알고리즘들이 부호화시 인접 시점의 영상들도 참조 화면으로 이용하고 있다. 그러나 HHI의 보고서[4]에 따르면 공간적인 참조의 경우 시간적인 참조보다 효율이 많이 떨어짐을 알 수 있다. 이는 동일 시간의 영상이라고 해도 카메라의 위치에 따라 조명과 물체의 배치 등이 조금씩 달라지기 때문이다.

그러나 이보다 더 큰 문제점이 있다. 현재 다시점 부호화의 실험

영상은 640×480 이상의 크기를 갖는 영상들이고, 현재 실험 영상의 시점은 카메라의 배치에 따라 다르지만 적게는 5대에서 많게는 100대의 카메라를 이용한다. 이에 따른 카메라의 간격은 좁게는 5cm에서 넓게는 20cm를 넘는다. 이는 실제 화면에서는 그 차이가 30 화소 이상의 차이로 반영된다. 그림 4는 640×480 크기의 “Exit” 영상에 대한 화면 전체 변이를 보여주고 있다.

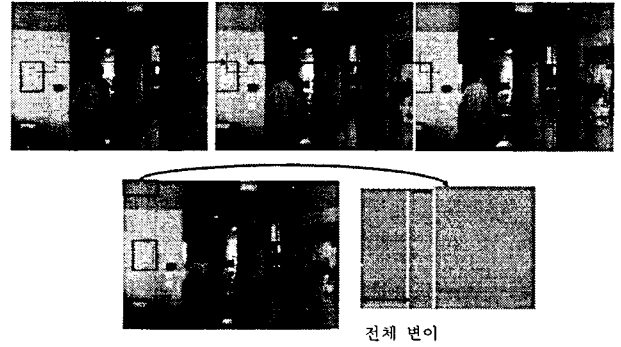


그림 4. 시점에 따른 전체 변이

그림 4에 보인 것처럼, 시점의 변화에 따른 전체 변이는 30 화소 이상이다. 따라서 공간적인 참조를 할 경우 기존의 탐색 범위인 ±32로는 올바른 움직임 예측이 불가능하다. 따라서 올바른 공간적인 참조를 위해서 제안하는 가변적 탐색 범위 알고리즘은 공간적인 참조인 경우에 대해서는 탐색 범위를 확장하여 ±64의 탐색 범위를 이용한다.

제안한 알고리즘은 격자 GOP 배치를 통해 고화질의 I 화면을 다시점 영상 전체에 효율적으로 분배하였고, 피라미드 GOP 구조를 통해 한 GOP에 대한 부호화 효율을 향상시켰다. 또한 공간적인 참조에 대해서는 가변적인 탐색 범위 알고리즘을 이용하였다. 그림 5는 제안하는 알고리즘의 부호화 구조도와 부호화 순서를 보여주고 있다.

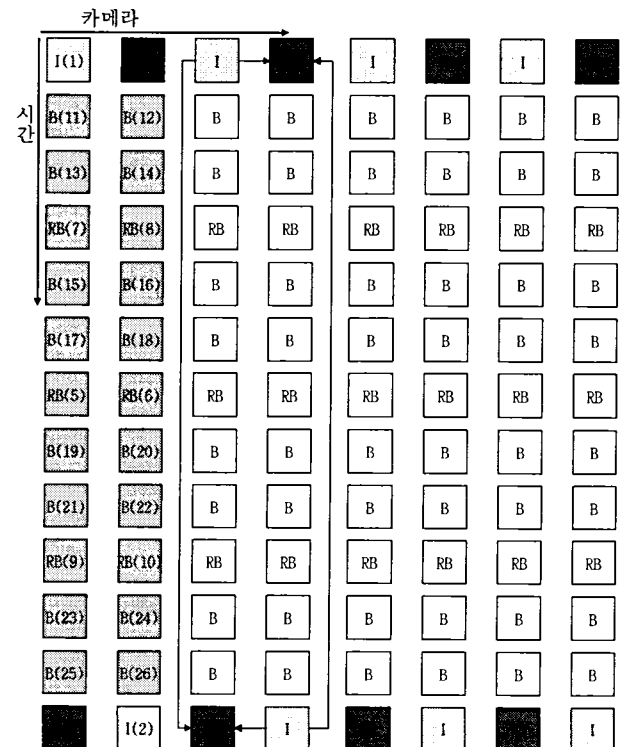


그림 5. 제안하는 방법의 부호화 구조도

4. 실험 결과 및 분석

본 논문에서 제안한 알고리즘을 검증하기 위해 MERL의 실험 영상(640×480, 8시점) “Exit”와 “Ballroom” 영상을 이용했다. 8시점 가운데 카메라의 위치상 중앙에 위치한 3, 4, 5, 6 시점의 250프레임에 대해 Anchor의 결과와 비교하였다. 비트율은 MPEG에서 권고한 비트율을 이용했고, 제안하는 알고리즘의 경우 피라미드 구조의 각 계층에 대한 중요도에 따라 수동적으로 QP 값을 조절했다. 표 1은 ‘Anchor’와 제안한 알고리즘의 부호화 효율을 비교하여 보여주고 있다 [5].

표 1. Anchor 방법과 제안한 알고리즘의 부호화 효율 비교

실험영상	비트율 (kbps)		PSNR (dB)	
	Anchor	제안한 알고리즘	Anchor	제안한 알고리즘
Exit	192.26	192.43	36.07	36.68
	256.32	255.21	37.05	37.57
	384.55	382.22	37.98	38.67
Ballroom	256.68	256.90	30.92	31.72
	384.74	383.16	32.88	33.48
	512.71	514.30	34.23	34.72

표 1에서 알 수 있듯이, 본 논문에서 제안한 알고리즘은 Anchor에 비해 비슷한 비트율에서 0.5~0.8 dB 정도 좋은 화질을 보였다. 그림 6은 각 실험 영상에 대한 비트율-왜곡 곡선을 보여주고 있다. 그림 6에서 보듯이, 저비트율과 고비트율 영역 모두에서 안정된 성능을 보인다.

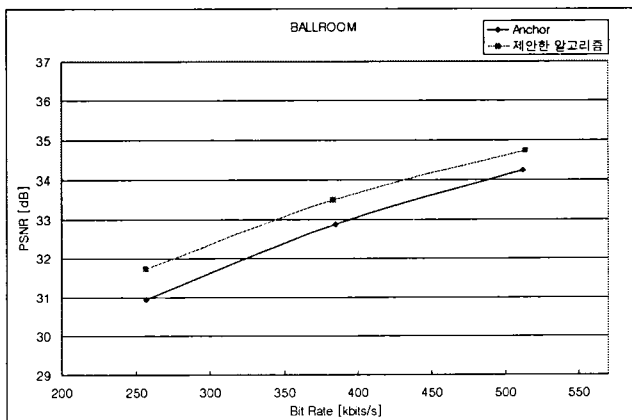
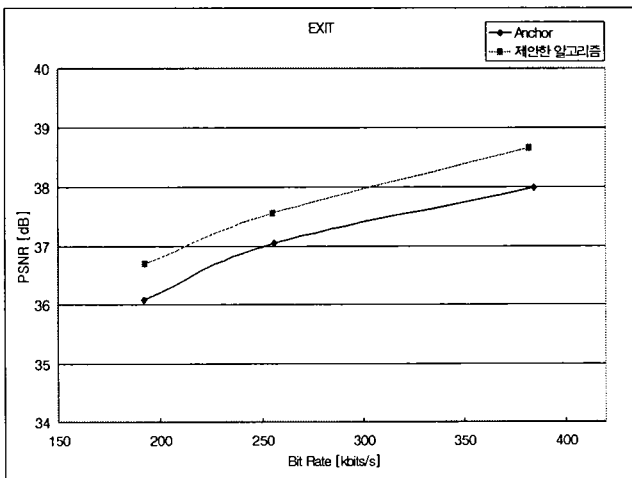


그림 6. 비트율-왜곡 곡선 (상: Exit, 하: Ballroom)

5. 결론

본 논문에서는 격자 피라미드 GOP 기반의 다시점 비디오 부호화 방법을 제안했는데, 이는 I 화면의 효과적인 배치를 위한 격자 GOP 배치와 각 GOP의 효율적인 부호화를 위한 피라미드 GOP 구조, 그리고 공간적인 참조에 대한 가변적 탐색 범위 알고리즘으로 구분된다. 격자 GOP 배치 알고리즘에서는 Anchor 방법에 비해 2배의 I 화면 간격을 이용하여 비트율을 줄이면서도 부호화 효율을 향상시킬 수 있었고, 피라미드 GOP 구조 알고리즘을 이용하여 각 GOP에 대한 부호화 효율의 향상과 시간적인 계위성에 대한 문제를 해결하였다. 마지막으로, 가변적인 탐색 범위 알고리즘을 통해서 화면의 전체 변이로 인한 움직임 예측의 비효율성을 극복하였다. 컴퓨터 모의실험을 통해 본 논문에서 제안한 다시점 부호화 방법이 Anchor 방법에 비해 동일 비트율에서 0.5~0.8 dB 정도의 개선된 화질을 보임을 확인하였다.

감사의 글

본 연구는 광주과학기술원(GIST) 실감방송연구센터(RBRC)를 통한 정보통신부 및 정보통신연구진흥원의 대학IT연구센터(ITRC) 지원 사업, 교육인적자원부 두뇌한국21 (BK21) 정보기술사업, 그리고 전자부품연구원(KETI)의 지원에 의한 것입니다.

참고 문헌

- [1] ISO/IEC JTC1/SC29/WG11 N6909, “Survey of Algorithms used for Multi-view Video Coding (MVC),” January 2005.
- [2] ISO/IEC JTC1/SC29/WG11 N7071, “Requirements on Multi-view Video Coding v.3,” April 2005.
- [3] ITU-T SG 16 Q.6 VCEG-N19, “H.264/MPEG-4 AVC Reference Software Enhancements,” January 2005.
- [4] ISO/IEC JTC1/SC29/WG11 M12301, “Statistical Evaluation of Spatiotemporal Prediction for MVC,” July 2005.
- [5] ISO/IEC JTC1/SC29/WG11 N7327, “Call for Proposals on Multi-view Video Coding,” July 2005.