

다시점 비디오 부호화를 위한 시간적 예측 구조

윤호순[†], 김미영^{††}

요 약

다시점 비디오는 3차원 정보를 표현하기 위한 영상으로 하나의 3차원 장면을 여러 시점에서 다수의 카메라로 촬영한 동영상이다. 영상들 사이에 존재하는 시간적 상관성과 화면간 상관성을 이용하는 다시점 비디오 부호화는 카메라의 수에 비례하여 데이터의 양이 늘어나기 때문에 계산량을 줄일 수 있는 다시점 비디오 부호화 기술이 필요하다. 본 논문에서는 다시점 비디오의 부호화 성능을 향상시키기 위한 효율적인 예측구조를 제안한다. 제안한 예측 구조는 다시점 비디오의 부호화 효율을 높이기 위하여 부호화되는 현재 화면과 현재 화면이 참조하는 참조 화면들과의 평균 거리, B계층 최대 인덱스 그리고 각 Bi 계층의 화면 수를 고려하였다. 제안한 예측 구조의 성능을 참조 예측 구조의 성능과 비교하였을 때 영상 화질 면에 있어서 제안한 예측 구조가 Fraunhofer-HHI의 계층적 B화면 구조보다 약 0.07~0.13 (dB) 성능 향상을 보였다. 발생하는 평균 초당 비트량에 있어서 제안한 예측 구조가 Fraunhofer-HHI의 계층적 B화면 구조보다 최대 6.5(Kbps) 감소하였다.

Temporal Prediction Structure for Multi-view Video Coding

Hyo-sun Yoon[†], Mi-Young Kim^{††}

ABSTRACT

Multi-view video is obtained by capturing one three-dimensional scene with many cameras at different positions. Multi-view video coding exploits inter-view correlations among pictures of neighboring views and temporal correlations among pictures of the same view. Multi-view video coding which uses many cameras requires a method to reduce the computational complexity. In this paper, we proposed an efficient prediction structure to improve performance of multi-view video coding. The proposed prediction structure exploits an average distance between the current picture and its reference pictures. The proposed prediction structure divides every GOP into several small groups to decide the maximum index of hierarchical B layer and the number of pictures of each B layer. Experimental results show that the proposed prediction structure shows good performance in image quality and bit-rates. When compared to the performance of hierarchical B pictures of Fraunhofer-HHI, the proposed prediction structure achieved 0.07~0.13 (dB) of PSNR gain and was down by 6.5(Kbps) in bitrate.

Key words: multi-view video(다시점 비디오), video compression(비디오 압축), prediction structure(예측 구조), hierarchical B layer(계층적 B 계층)

※ 교신저자(Corresponding Author): 김미영, 주소: 전라남도 담양군 담양읍 향교리 262번지(517-802), 전화: 061) 380-8651, FAX: 061)380-8628, E-mail: kimmee@dorip.ac.kr
접수일: 2011년 3월 29일, 수정일: 2011년 6월 21일
완료일: 2012년 8월 19일

[†] 정회원, 전남도립대학교 보건의료과
(E-mail: estheryoon@ hotmail.com)

^{††} 정회원, 전남도립대학교 보건의료과

※ 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 연구임 (No. 2011-0026947).

1. 서 론

데이터 압축 기술과 데이터 전송 속도 향상 그리고 통신 기술의 발달로 다양한 멀티미디어 콘텐츠를 사용자에게 제공할 수 있게 되었다. 그리고 자유 시점 TV (free viewpoint television), 3차원 TV와 같은 멀티미디어 시스템의 발달로 인하여 실감 미디어 콘텐츠에 대한 수요가 급증하고 있다[1]. 이에 따라 3차원 비디오의 중요성이 커지고 있으며 다시점 비디오는 3차원 영상에 대한 다양한 사용자의 요구를 충족시킬 수 있는 대안으로 주목받고 있다.

다시점 비디오는 여러 개의 카메라로 동시에 하나의 3차원 장면을 촬영한 동영상의 집합으로 다시점 비디오는 사용자에게 임의 시점을 제공하며 여러 시점의 영상을 합성하여 보다 넓은 화면을 제공할 수 있다. 그리고 다시점 영상 디스플레이 장치를 통해 사용자에게 입체감 있는 3차원 영상을 제공하지만 많은 기억 공간과 대역폭을 필요로 한다[2]. 다시점 비디오는 카메라의 수에 비례하여 데이터의 양이 늘어나기 때문에 다시점 영상 정보를 효율적으로 부호화하는 기술이 필요하다. 그래서 다시점 비디오 부호화(Multi-view Video Coding: MVC) 기술에 대한 표준화 작업이 시작되었고 최근에 다시점 비디오를 효과적으로 부호화 방법을 국제 표준으로 제정하였다[3,4].

다시점 비디오 부호화의 초기 단계에서 이웃하는 시점 사이에 존재하는 화면간의 상관성을 이용하지 않아서 부호화 성능이 좋지 않았다. 부호화 성능을 향상시키기 위하여 시간적 상관성뿐만 아니라 화면간의 상관성을 이용한 예측 구조들이 제안되었다. 다시점 비디오 부호화를 위한 예측 구조들은 주로 영상들 사이에 존재하는 시간적 상관성과 화면간의 상관성을 이용하여 영상들 사이에 존재하는 중복된 데이터를 제거함으로써 부호화 성능을 향상시켰다. 시간적 상관성과 화면간 상관성을 이용하여 움직임 추정과 변이 추정을 수행하는 계층적 B화면 구조가 다시점 비디오 부호화를 위하여 사용되고 있다[5-8].

계층적 B화면 구조의 성능을 향상시키기 위하여 여러 예측 구조들이 제안되었다. Park의 시간적 상관성을 이용한 예측 구조는 현재 화면과 참조하는 화면 사이의 평균거리를 줄였고 화면간의 예측 구조는 I 시점과 P 시점의 위치를 결정하기 위하여 글로벌 변

이를 이용하였다[9]. Herenlong의 제안 기법은 시간적 상관성이 강한 영상, 공간적 상관성이 강한 영상, 그리고 시간적 상관성과 공간적 상관성이 거의 같은 영상에 대하여 각각 적응적으로 참조 화면 모드를 결정하였다[10]. 그러나 이 기법은 여러 모드들을 사용하므로 많은 계산량을 요구한다. Feng은 상관성 분석을 기반으로 다양한 예측 구조들을 제안하였다[11]. 즉, 시간적 상관성이 우세한 영상에서 사용되는 예측 구조, 공간적 상관성이 우세한 영상에서 사용되는 예측 구조, 그리고 시간적 상관성과 공간적 상관성이 우세한 영상에서 사용되는 예측 구조들을 제안하였다.

본 논문에서 제안하는 예측 구조는 다시점 비디오의 부호화 효율을 높이기 위하여 부호화되는 현재 화면과 현재 화면이 참조하는 참조 화면들과의 평균 거리와 B계층의 최대 인덱스와 각 B계층의 화면 수를 고려하였다.

본 논문의 구성은 2장에서 다시점 비디오 부호를 위한 Fraunhofer-HHI의 계층적 B화면 구조와 기존 방법들을 설명하고, 3장에서 제안한 예측 구조를 기술한다. 그리고 4장에서 제안한 예측 구조 성능을 기술한 후, 5장에서 결론으로 맺는다.

2. Fraunhofer-HHI의 계층적 B화면 구조와 기존 방법들

다시점 비디오는 하나의 3차원 장면을 동시에 여러 대의 카메라로 촬영한 영상들의 집합이므로 영상들 사이에 중복성이 존재한다. 즉, 인접한 시점의 영상들 사이에는 화면간 중복성과 같은 시점 영상들 사이에는 시간적 중복성이 존재한다[12]. 그러므로 다시점 비디오 부호화는 영상들 사이에 존재하는 시간적 상관성과 화면간 상관성을 이용하여 시간적으로 중복된 데이터와 시점간 중복된 데이터를 제거함으로써 20%~50%의 압축효율을 제공한다[13]. 다시점 비디오 시스템에서는 다시점 비디오를 부호화하기 위해 그림 1의 Fraunhofer-HHI의 계층적 B화면 구조를 사용한다. 즉, 그림 1의 Fraunhofer-HHI의 계층적 B화면 구조가 성능 평가를 위한 구조로 사용되고 있다. 본 논문에서는 Fraunhofer-HHI의 예측 구조를 표준 예측 구조로 기술한다.

그림 1에서 S_n 은 n 번째 시점의 카메라를 의미하

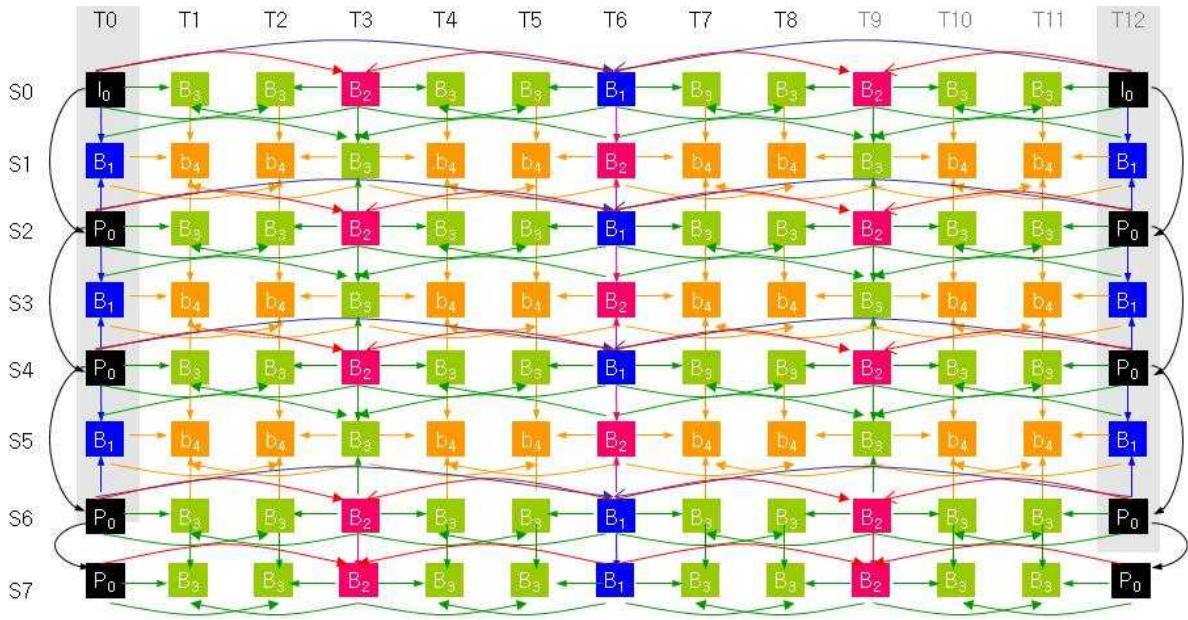


그림 1. Fraunhofer-HHI의 계층적 B화면 구조

고 T_n 은 시간적으로 n 번째 화면을 나타낸다. 화살표는 이웃하는 화면들 사이의 예측 참조 관계를 나타낸다. 기존 시스템과의 호환성을 유지하기 위하여 다른 시점과 상관없이 독립적으로 복원할 수 있는 시점을 I시점, 부호화가 완료된 하나의 시점만을 참조하여 예측 부호화하는 시점을 P시점 그리고 인접해 있는 두 개의 시점을 참조하여 예측 부호화하는 시점을 B시점이라고 한다. 그림 1의 S_0 은 I시점 그리고 S_2, S_4, S_6, S_7 들은 P 시점, S_1, S_3, S_5 들은 B시점이다. 그림 1의 표준 예측 구조는 시간 예측을 위하여 계층적 B구조(hierarchical B picture structure)를 사용하며, $S_0 \sim S_7$ 시점에서 각 GOP의 첫 화면인 T_0 과 T_{12} 는 앵커(anchor) 화면들로 시간적 임의접근과 에러 전파 방지를 위하여 0.5 또는 1초 간격으로 삽입된다 [3,4].

표준 예측 구조의 계산량을 줄이기 위하여 여러 가지 예측 구조들이 제안되었다[9-11]. 기존 예측 구조들은 시간적 상관성을 이용하거나 화면간 상관성을 이용하거나 시간적 상관성과 화면간 상관성을 모두 이용한 예측 구조들을 제안하였으며 현재 화면과 현재 화면이 참조하는 참조 화면들 사이의 평균거리 이용한 예측 구조, I 시점의 위치를 변경시키는 예측 구조들을 제안하였다. 그리고 시간적 상관성이 강한 영상, 공간적 상관성이 강한 영상, 그리고 시간적 상관성과 공간적 상관성이 거의 비슷한 영상에 대하여

각각 적응적으로 예측 구조를 결정하는 기법들을 제안하였다.

이와 같은 기존 예측 구조들은 다시점 비디오를 부호화하기 위하여 사용할 예측 구조를 결정하거나 참조 화면 모드를 적응적으로 결정하기 위하여 부호화 대상의 영상이 시간적 상관성이 강한 영상인지 공간적 상관성이 강한 영상인지 시간적 상관성과 공간적 상관성이 비슷한 영상인지 먼저 분석하여야 한다. 그리고 I시점 위치를 결정하기 위해 이웃하는 시점들의 영상들 중 가장 많이 중첩되는 영상의 시점을 찾고, 이 시점을 I시점으로 결정하였다. 즉 예측 구조를 결정하기 위한 전처리 과정이 필요하다.

3. 제안한 예측구조

다시점 비디오의 부호화 효율을 높이기 위하여 본 논문에서 제안하는 예측 구조는 부호화되는 현재 화면과 현재 화면이 참조하는 참조 화면들과의 평균 거리뿐만 아니라 B계층의 최대 인덱스 그리고 각 B_i 계층에 속하는 화면 수를 고려하였다. B_i 에서 i 를 인덱스라 정의한다. 일반적으로 B계층 최대 인덱스는 GOP를 몇 개의 그룹으로 분할하느냐에 따라 달라진다.

그림 2에서 n 은 길이가 12인 GOP를 여러 개의 작은 그룹으로 분할할 때 생성되는 그룹의 수로 B1 화

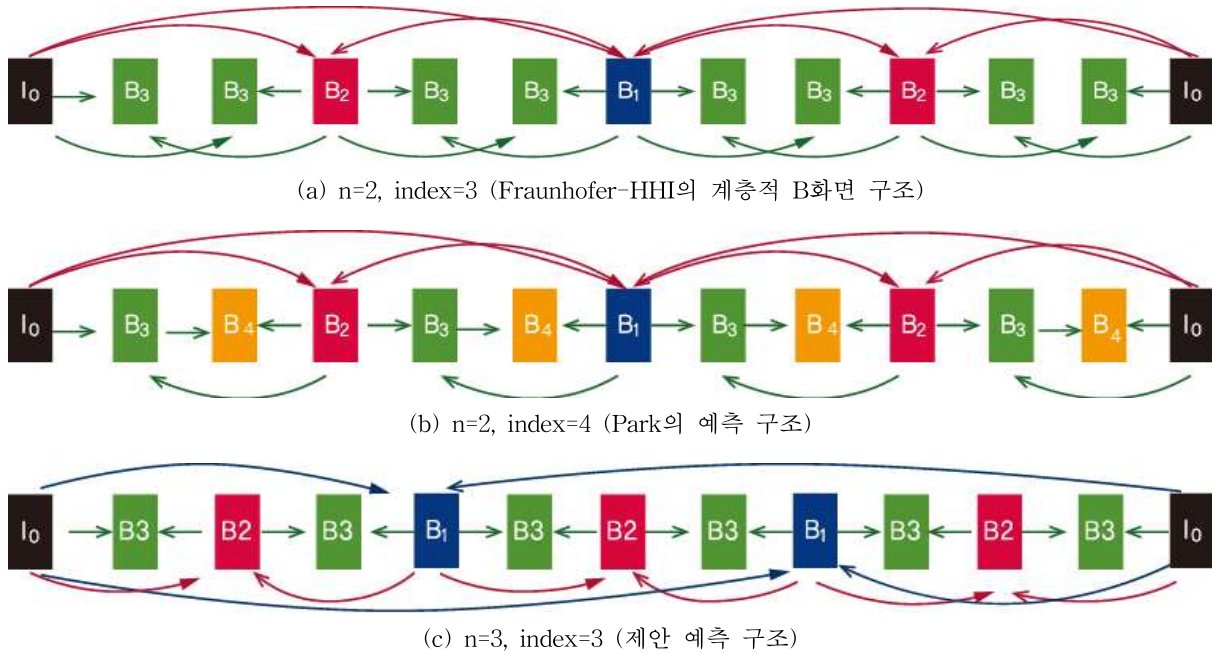


그림 2. GOP 분할 그룹 수와 B계층 인덱스

면 수에 1을 더한 것과 같다. 그리고 B_i 에서 i 를 인덱스로 정의하였을 때, index는 B계층 최대 인덱스이다. 길이가 12인 GOP가 만들 수 있는 B계층의 최대 인덱스는 아래와 같다.

$$\text{Layer_max} = \lceil \log_n \text{GOPlength} \rceil \quad (1)$$

식 (1)에서 GOPlength 는 GOP의 길이를 나타내고, n 은 길이가 12인 GOP의 분할 그룹의 수를 나타낸다[9]. 즉, n 은 하나의 GOP를 몇 개의 그룹으로 나눌 것인가를 나타낸다. 그림 2(a)와 그림 2(b)에서 길이가 12인 GOP를 2개 그룹으로 나누었을 때 생성되는 Layer_max 는 4이므로 B계층 최대 인덱스는 1에서 4사이의 정수이고, 그림 2(c)에서 GOP를 3개 그룹으로 나누었을 때 생성되는 Layer_max 는 3이므로 B계층 최대 인덱스는 1에서 3사이의 정수이다. 즉, 그림 2(a)와 그림 2(b)에서 GOP가 12일 때 하나의 GOP를 2개 그룹으로 나누었을 때 최대 B_4 계층까지 생성될 수 있고, 그림 2(c)와 같이 GOP를 3개 그룹으로 나누었을 때 B_3 계층까지 생성될 수 있다.

길이가 12인 GOP를 여러 그룹으로 나눌 때, 생성되는 그룹 수와 각 B_i 계층의 화면 수 그리고 부호화되는 현재 화면과 현재 화면이 참조하는 참조 화면들과의 평균거리를 표 1에 나타내었다.

B계층 최대 인덱스가 작을수록 발생하는 비트량은 많아지고 영상의 화질이 좋다. 반면에 B계층 최

표 1. 분할 그룹 수와 각 B계층의 화면 수와 평균 거리

	GOP 12 분할 그룹 수		
	그림 2(a)	그림 2(b)	그림 2(c)
	n=2	n=2	n=3
B_1 계층	1개	1개	2개
B_2 계층	2개	2개	3개
B_3 계층	8개	4개	6개
B_4 계층		4개	
평균거리	4.36	4	4.36

대 인덱스가 클수록 발생하는 비트량은 적어지며 영상 화질도 좋지 않다. 그러므로 영상 화질과 발생하는 비트량을 고려하여 각 GOP에서 B계층의 최대 인덱스와 각 B계층의 화면 수를 결정하여야 한다.

영상 화질과 발생하는 비트량을 고려하여 각 GOP에서 B계층의 최대 인덱스와 각 B계층의 화면 수를 결정하기 위하여 부호화되는 현재 화면과 현재 화면이 참조하는 참조화면들과의 평균거리를 이용한다. 분할되는 그룹 수, B계층 최대 인덱스 그리고 B_i 계층의 화면 수에 따라 GOP에서 화면들을 부호화 할 때 부호화되는 현재 화면과 현재 화면이 참조하는 참조화면들과의 거리가 달라진다. 부호화되는 현재 화면과 현재 화면이 참조하는 참조화면들과의 거리를 구하기 위하여 식 (2)를 이용한다[9].

$$distance_i = \frac{GOPlength}{N^{i-1}} \quad (\text{단, } distance < 2 \text{ 이면, } distance = 2) \quad (2)$$

$$Distance_s = \sum_{i=1}^{\max} i_{Number} \times distance_i \quad (3)$$

$$Distance_{mean} = \frac{Distance_s}{GOPlength-1} \quad (4)$$

식 (2)에서 N은 분할되는 그룹 수이고, GOPlength는 GOP 길이이며 i는 B계층 인덱스이다. 그림 2(a)에서 B₁계층의 distance₁는 $\frac{12}{2^{1-1}}$ 이고 B₂계층의 distance₂는 $\frac{12}{2^{2-1}}$ 이며 B₃계층의 distance₃는 $\frac{12}{2^{3-1}}$ 이다. 식 (2)에서 B_i계층의 참조화면들과의 거리인 distance_i가 2미만이면 부호화되는 현재 화면과 현재 화면이 참조하는 참조화면들과의 거리를 2로 한다. B_i계층의 distance_i가 2미만인 경우에도 부호화되는 현재 화면이 참조하는 화면들은 순방향과 역방향에서 각각 가장 가까운 화면이기 때문이다. 분할 그룹 수 N은 [2, GOP_length]사이의 값을 가질 수 있는데 N이 커질수록 GOP에서 부호화되는 현재 화면과 현재 화면이 참조하는 참조화면들과의 평균 거리가 길어진다. 식 (3)에서 max는 GOP를 분할하였을 때 생성되는 B계층 최대 인덱스를 의미한다. 예를 들면, 그림 2(a)에서 max는 3이고 그림 2(b)에서 max는 4이며 그림 2(c)에서 max는 3이다. 그리고 식 (3)에서 i_{Number}는 B_i계층의 화면 수를 의미하고 distance_i는 현재 화면과 현재 화면이 참조하는 참조화면들과의 거리이다 [9].

그림 2(a)는 GOP를 두 그룹으로 분할한 구조로 분할되는 그룹 수 N=2이고 max=3이며 생성되는 B₁계층의 화면 수는 1, B₂계층의 화면 수는 2, B₃계층의 화면 수는 8이다. 그리고 distance₁=12, distance₂=6, distance₃=3이다. 그림 2(b)는 GOP를 두 그룹으로 분할한 구조로 분할되는 그룹 수 N=2이고 max=4이며 생성되는 B₁계층의 화면 수는 1, B₂계층의 화면 수는 2, B₃계층의 화면 수는 4, B₄계층의 화면 수는 4이다. 그림 2(c)에서 N=3이고 max는 3이며 생성되는 B₁계층의 화면 수는 2, B₂계층의 화면 수는 3, B₃계층의 화면 수는 6이다. 식 (3)을 이용하여 각 예측 구조의 현재 화면과 현재 화면이 참조하는 참조화면들과의 거리의 합을 구하면 그림 2(a)의 경우 Distance_s는 1*12 + 2*6 + 8*3이므로 48이다. 식 (4)을 이용하여 각 예측 구조의 현재 화면과 참조화면들과

의 평균거리를 구한다. 그림 2(a)의 경우 Distance_{mean}는 $\frac{48}{12-1}$ 이므로 4.36이고 그림 2(b)의 Distance_{mean}는 4, 그림 2(c)의 Distance_{mean}는 4.36이다.

본 논문에서 다시점 비디오의 부호화 효율을 높이기 위하여 부호화하는 현재 화면과 현재 화면이 참조하는 참조화면들과의 평균 거리뿐만 아니라 B계층 인덱스와 그리고 각 B_i계층에 속하는 화면 수를 고려하였다. 영상 화질과 발생하는 비트량을 고려하여 각 GOP에서 B계층의 최대 인덱스와 각 B계층의 화면 수를 결정하기 위하여 Fraunhofer-HHI의 계층적 B화면 구조의 최대 인덱스와 각 B계층의 화면 수 그리고 평균 거리를 기준으로 하였다. B계층의 최대 인덱스와 각 B계층의 화면 수 그리고 평균 거리는 영상 화질과 발생하는 비트량에 영향을 주므로 최대 인덱스는 Fraunhofer-HHI의 계층적 B화면 구조와 같으면서 B계층의 최대 인덱스의 화면수가 Fraunhofer-HHI의 계층적 B화면 구조의 최대 인덱스의 화면수보다 적은 구조를 제안하였다. 일반적으로 최대 인덱스가 같은 경우 최대 인덱스의 화면수가 적을수록 영상의 화질이 좋다. 즉, B계층의 최대 인덱스를 제외한 계층에서의 화면수가 Fraunhofer-HHI의 계층적 B계층 구조의 화면수보다 많아야한다. 그리고 최대 인덱스가 Fraunhofer-HHI보다 작아지면 영상 화질은 좋아지지만 발생하는 비트량이 많아진다. 또한 B계층의 최대 인덱스는 평균거리에도 영향을 준다. 일반적으로 B계층의 최대 인덱스가 작아질수록 평균거리가 길어진다.

본 논문에서 제안하는 예측 구조인 그림 2(c)의 예측 구조를 Fraunhofer-HHI의 계층적 B화면 구조인 그림 2(a)와 비교하였을 때, B₁계층의 화면 수는 다르지만 평균 거리와 B계층 최대 인덱스는 같다. 제안하는 예측 구조의 B₁계층의 화면 수는 2, B₂계층의 화면 수는 3, B₃계층의 화면 수는 6이다. Fraunhofer-HHI의 계층적 B계층 구조인 경우, B₁계층, B₂계층, B₃계층의 화면 수는 각각 1, 2, 8이다. 제안하는 예측 구조와 Fraunhofer-HHI의 계층적 B화면 구조의 평균 거리가 같지만 제안하는 예측 구조의 B₁계층의 화면 수와 B₂계층의 화면 수가 참조 예측 구조의 B₁계층, B₂계층의 화면 수 보다 많으므로 제안하는 예측 구조가 Fraunhofer-HHI의 계층적 B화면 구조보다 영상 화질 면에서 좋고 비슷한 비트량을 보일 것

이다. 그리고 제안하는 예측 구조인 그림 2(c)의 예측 구조를 Park의 시간적 상관성을 이용한 예측 구조인 그림 2(b)와 비교하였을 때, 제안하는 예측 구조의 최대 인덱스가 Park의 시간적 상관성을 이용한 예측 구조의 최대 인덱스보다 작으므로 발생하는 비트량은 증가하는 반면에 영상 화질은 좋아질 것이다.

4. 실험 결과

제안하는 예측 구조의 성능을 확인하기 위하여 제안하는 예측 구조를 JMVM 6.0에서 구현하였다. 실험 영상으로 Exit, Ballroom, Uli를 사용하였다. 실험 조건 표 2에 나타내었다.

본 논문에서 그림 3의 제안한 예측 구조 성능과 그림 1의 Fraunhofer-HHI의 계층적 B화면 구조 그리고 Park의 시간적 상관성을 이용한 예측 구조의 성능을 표 3, 표 4, 표 5에 제시하였다. 표 3, 표 4, 표 5에 제시된 실험결과는 QP가 37이다. Exit의 실험 결과는 표 3, Uli의 실험 결과는 표 4, Ballroom의 실험 결과는 표 5에 나타냈다. 표 3에 제시된 Exit 영상의 실험 결과를 살펴보면 제안한 예측 구조가 Fraunhofer-HHI의 계층적인 B화면 구조보다 영상 화질 면에서 평균 0.07(dB) 성능 향상을 보였으며 발생 비트량에 있어서 평균 1.008(Kbps)의 감소를 보였다. 그리고 제안한 예측 구조의 성능과 Park의 시간적 상관성을 이용한 예측 구조의 성능을 비교하였

표 2. 실험 조건

영 상	영상의 크기	기본 QP	탐색 범위	프레임 수
Exit	640*480	22, 27, 32, 37	±96	100
Ballroom	640*480			
Uli	1024*768			

을 때, 제안한 예측 구조가 Park의 시간적 상관성을 이용한 예측 구조보다 영상 화질 면에서 평균 0.14 (dB) 성능 향상을 보였으며 비트량에 있어서 평균 3.86(Kbps)의 증가를 보였다.

표 4에 제시된 Uli 영상의 실험 결과를 살펴보면 제안한 예측 구조가 Fraunhofer-HHI의 계층적 B화면 구조보다 화질 면에서 평균 0.13(dB) 성능 향상을 보였으며 발생 비트량에 있어서 평균 6.776(Kbps) 감소를 보였다. 그리고 제안한 예측 구조의 성능과 Park의 시간적 상관성을 이용한 예측 구조의 성능을 비교하였을 때, 제안한 예측 구조가 Park의 시간적 상관성을 이용한 예측 구조보다 영상 화질 면에서 평균 0.22(dB) 성능 향상을 보였으며 발생 비트량에 있어서 평균 28.53(Kbps)의 증가를 보였다. 표 5에 제시된 Ballroom 영상의 실험 결과에서 제안한 예측 구조는 영상 화질 면에 있어서 Fraunhofer-HHI의 계층적 B화면 구조보다 평균 0.10(dB) 성능 향상을 보였지만 발생 비트량에 있어서 평균 3.116(Kbps) 증가를 보였다. 그리고 제안한 예측 구조의 성능과

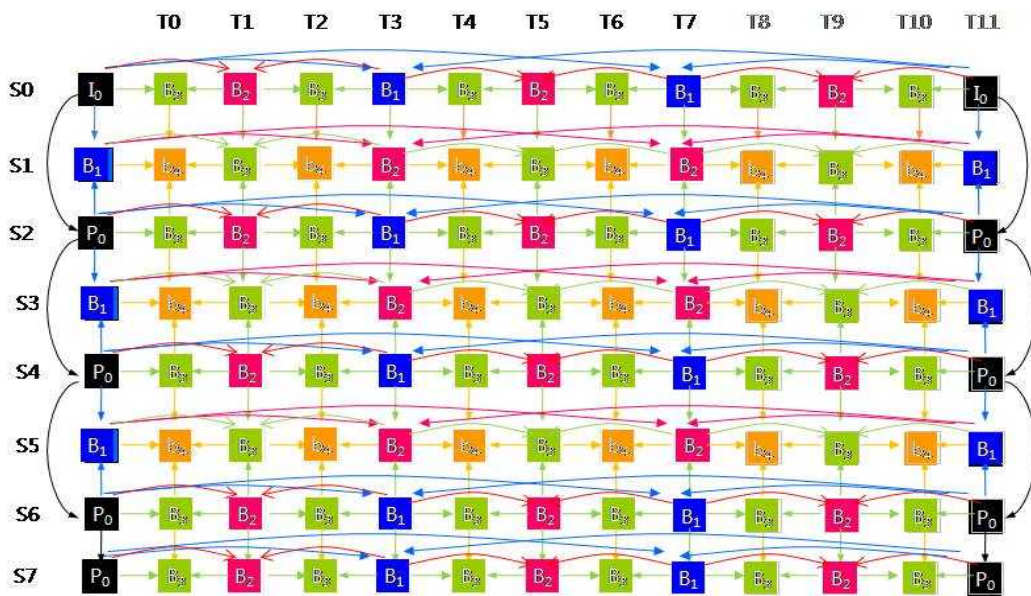


그림 3. 제안 예측 구조

표 3. Exit 실험 결과

카메라 번호	Fraunhofer-HHI의 계층적 B화면 구조		Park의 예측 구조		제안 예측 구조	
	PSNR (db)	bit rate (Kbps)	PSNR (db)	bit rate (Kbps)	PSNR (db)	bit rate (Kbps)
0	35.1402	113.536	35.0895	110.614	35.1984	113.058
1	34.0961	60.308	34.0086	56.506	34.1864	59.076
2	34.8471	95.966	34.7808	92.04	34.9181	95.8
3	33.7928	72.814	33.6904	67.922	33.864	70.764
4	34.2998	118.052	34.2226	113.02	34.3741	118.05
5	33.6858	89.168	33.5924	83.034	33.7472	87.45
6	33.7063	149.758	33.6319	143.964	33.7872	149.19
7	33.6582	131.69	33.5745	125.206	33.7121	129.834

표 4. Uli 실험 결과

카메라 번호	Fraunhofer-HHI의 계층적 B화면 구조		Park의 예측 구조		제안 예측 구조	
	PSNR (db)	bit rate (Kbps)	PSNR (db)	bit rate (Kbps)	PSNR (db)	bit rate (Kbps)
0	31.5464	942.772	31.4762	902.02	31.6826	934.506
1	29.1249	753.902	29.0192	702.984	29.2822	744.316
2	32.0423	798.246	31.9644	758.376	32.1832	792.51
3	30.22	592.552	30.0951	548.394	30.3479	581.81
4	32.95	609.114	32.878	578.476	33.0988	603.422
5	31.6668	417.384	31.5465	388.08	31.8147	409.764
6	33.1867	552.848	33.1191	526.736	33.3269	547.406
7	32.2128	675.804	32.1375	655.086	32.319	674.674

표 5. Ballroom 실험 결과

카메라 번호	Fraunhofer-HHI의 계층적 B화면 구조		Park의 예측 구조		제안 예측 구조	
	PSNR (db)	bit rate (Kbps)	PSNR (db)	bit rate (Kbps)	PSNR (db)	bit rate (Kbps)
0	32.1016	295.704	32.0472	286.62	32.1929	298.32
1	31.2187	156.412	31.14	146.986	31.3004	158.25
2	32.157	245.95	32.0996	236.526	32.267	250.278
3	31.2049	152.912	31.1262	144.846	31.2941	154.648
4	31.5052	268.79	31.4518	258.59	31.6076	273.244
5	31.7143	172.128	31.6315	31.4518	31.7869	172.64
6	31.8597	265.312	31.8088	254.918	31.966	270.094
7	31.1416	245.28	31.0823	234.33	31.239	249.946

표 6. BD-PSNR와 BD-BitRate

	Fraunhofer-HHI의 계층적 B화면 구조			제안예측구조		BD_PSNR_Y(dB)	BD-BitRate(%)
	QP	BitRate	PSNR_Y(dB)	BitRate	PSNR_Y(dB)		
EXIT	22	791.40175	39.854238	810.1025	39.928638	0.05207	-1.7714
	27	335.2205	38.306213	340.826	38.383063		
	32	175.73675	36.435963	177.03825	36.512163		
	37	103.9915	34.15205	102.90275	34.223438		
	Fraunhofer-HHI의 계층적 B화면 구조			제안예측구조			
	QP	BitRate	PSNR_Y(dB)	BitRate	PSNR_Y(dB)	BD_PSNR_Y(dB)	BD-BitRate(%)
ULI	22	4148.4543	38.98629	4073.7821	39.24959	0.309421	-7.35393
	27	2172.191	36.9371	2133.0916	37.1671		
	32	1210.34	34.3968	1192.1849	34.6468		
	37	667.82775	31.61864	661.0485	31.88194		
	Fraunhofer-HHI의 계층적 B화면 구조			제안예측구조			
	QP	BitRate	PSNR_Y(dB)	BitRate	PSNR_Y(dB)	BD_PSNR_Y(dB)	BD-BitRate(%)
BALL ROOM	22	1491.287	39.05565	1515.0608	39.147835	0.05159	-1.29959
	27	748.254	36.81616	758.9575	36.91965		
	32	399.8633	34.25924	405.9877	34.380438		
	37	225.6924	31.61254	227.9725	31.710488		

Park의 시간적 상관성을 이용한 예측 구조의 성능을 비교하였을 때, 제안한 예측 구조가 Park의 시간적 상관성을 이용한 예측 구조보다 영상 화질 면에서 평균 0.15(dB) 성능 향상을 보였으며 발생 비트량에 있어서 평균 11.792(Kbps)의 증가를 보였다.

표 6에 제안하는 예측 구조와 Fraunhofer-HHI의 계층적 B화면 구조의 실험결과인 PSNR과 BitRate을 비교하였을 때 BD-PSNR와 BD-BitRate를 제시하였다. 영상의 화질 면에서 제안 예측 구조가 Fraunhofer-HHI의 계층적 B화면 구조보다 0.05~0.3(dB) 화질 향상을 보였으며 발생하는 비트량에 있어서 1.774~7.35% 비트량 감소를 보였다.

5. 결 론

본 논문에서는 다시점 비디오의 부호화 효율을 높이기 위하여, 즉 영상 화질을 향상시키면서 발생하는 비트량을 줄이기 위한 예측 구조를 제안하였다. 제안한 예측 구조는 부호화되는 현재 화면과 현재 화면이 참조하는 참조 화면들과의 평균거리뿐만 아니라 B

계층 최대 인덱스 그리고 각 B_i 계층의 화면 수를 고려하였다. 제안하는 예측 구조와 Fraunhofer-HHI의 계층적 B계층 구조는 현재 화면과 현재 화면이 참조하는 참조 화면들과의 평균 거리와 최대 인덱스는 같지만 제안하는 예측 구조의 B_1 계층의 화면 수와 B_2 계층의 화면 수가 다르다. 화질 면에서 제안하는 예측 구조가 Fraunhofer-HHI의 계층적 B화면 구조보다 약 0.05~0.3(dB) 성능 향상을 보였고 초당 비트량에 있어서 1.774 ~7.35% 비트량 감소를 보였다. 본 논문에서 제안한 예측구조는 시간적 구조에서의 부호화 순서를 변경하였는데 향후 I시점과 P시점의 위치를 변화시키는 화면간의 예측 구조를 추가하면 제안한 예측 구조보다 더 좋은 성능을 보일 것이다.

참 고 문 헌

[1] 호요성, 이천, “3D 비디오 부호화를 위한 표준화 동향과 요소 기술,” 한국멀티미디어학회지, 제14권, 제2호, pp. 24-36, 2010.
 [2] ITU-T Rec. & ISO/IEC 14496-10 AVC,

- Advanced Video Coding for Generic Audio Visual services*, 2005.
- [3] ISO/IEC JTC1/SC29/WG11 N10357, *Vision on 3D Video*, 2009.
- [4] A. Vetro, P. Pandic, H. Kimata, A. Smolic, and Y. Wang, *Joint Multi-view Video Model (JMVM) 8.0*, Joint Video Team of ISO/IEC JTC21/SC29/WG11 and ITU-T SG16 Q.6, JVT-AA207, Switzerland, 2008.
- [5] P. Merkle, K. Muller, A. Smolic, and T. Wiegand, "Efficient Compression of Multi-view Video Exploiting Inter-view Dependencies Based on H.264/MPEG4-AVC," *Proc. of IEEE International Conference on Multimedia and Exposition*, pp. 1717-1720, 2006.
- [6] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding," *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 17, No.11, pp. 1461-1473, Nov. 2007.
- [7] ISO/IEC JTC1/SC29/WG11, *Joint Multiview Video Model (JMVM) 8*, Doc. N9762, 2008.
- [8] B. Zatt, M. Shafique, S. Bampi, and J. Henkel, "A Multi-Level Dynamic Complexity Reduction Scheme for Multiview Video Coding," *IEEE International Conference on Image Processing (ICIP)*, pp. 761-764, 2011.
- [9] P.K. Park, K.J. Oh, and Y.S. Ho, "Efficient View-temporal Prediction Structures for Multi-view Video Coding," *Electronics Letters*, Vol.44, No.2, pp. 102-103, 2008.
- [10] R.L. He, "A Multiview Video Coding Method with Adaptive Selection of Reference Frame Modes," *Journal of Computer-aided Design And Computer Graphics*, Vol.30, No.12, pp. 2205-2211, 2007.
- [11] F. Lu, P. An, Z. Zhang, and L. Shen, "Multi-view Video Coding Based on Sequence Correlation," *Audio Language and Image Processing (ICALIP)*, pp. 1227-1232, 2010.
- [12] 최정현; 이경환, "움직임의 시간적 연속성을 이용한 적응적 움직임 추정 알고리즘," 한국멀티미디어학회논문지, v.7, no.8, pp.1025-1034, 2004년 8월
- [13] P. Merkle, "Efficient Prediction Structures for Multiview Video Coding," *IEEE TCSTVT*, Vol. 17, No. 11, pp. 1461-1473, 2007.

김 미 영



1994년 3월~1997년 8월 전남대학교 전산학과 이학 박사
 1998년 3월~현재 전남도립대학교 보건의료과 교수
 관심분야 : 영상처리, 영상부호화, 멀티미디어 시스템

윤 호 순



2002년 3월~2005년 2월 전남대학교 전산학과 이학 박사
 관심분야 : 영상처리, 영상부호화, 멀티미디어 시스템