JMVC에서의 효율적인 예측구조

김미영*, 윤효순**
*전남도립대학교 보건의료학과
*전남대학교 전산학과
e-mail:estheryoon@hotmail.come, imkimmee@naver.com

Efficient Prediction Structure on Joint Multi-view Video Coding

Mi-Young Kim*, Hyo-Sun Kim**
*Dept of Health Medical, Jeonnam Provincial College
**Dept of Computer Science, Chonnam National University

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 B계층의 화면 수를 고려하였다. 제안한 예측 구조의 성능을 참조 예측 구조의 성능과 비교하였을 때 영상 화질 면에 있어서 제안한 예측 구조가 참조 예측 구조보다 약 0.07~0.13 (dB) 성능 향상을 보였다. 발생되는 평균 초당 비트량에 있어서 제안한 예측 구조가 참조 예측 구조보다약 +3 ~ - 6.5(Kbps) 감소하였다.

1. 서론

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 제공한다. 그러나 다시점 비디오는 카메라의 수에 비례하여 데이터의 양이 늘어나기 때문에 다시점 영상 정보를 효율적으로 부호화하는 기술이 필요하다. 이를 위하여 다시점 비디오 부호화 (Multi-view Video Coding: MVC) 방법을 국제 표준으로 제정하였다[1][2].

다시점 비디오 부호와의 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 제거함으로 부호화 성능을 향상시켰다. 다시점 비디오를 부호하기 위하여 시간적 상관성과 화면간 상관성을 이용한 계층적 B화면 구조가 사용되고 있다[3][4][5].

계층적 B화면 구조-참조 예측 구조-의 성능을 향상시키기 위하여 여러 예측 구조들이 제안되었다. Park의 시간적 상관성을 이용한 예측 구조는 현재 화면과 참조하는 화면사이의 평균거리를 줄였고 화면간의 예측 구조는 I 시점과 P 시점의 위치를 결정하기 위하여 글로벌 변이를 이용하였다[6]. 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[7]. 그러나 이 기법은 여러 모드들를 사용하므로 많은 계산량을 요구한다. Feng은 상관성 분석을 기반으로 다양한 예측 구조들을 제안하였다[8]. 즉, 시간적 상관성이 우세한 영상에서 사용되는 예측 구조, 시간적 상관성이 우세한 영상에서 사용되는 예측 구조, 그리고 시간적 상관성과 공간적 상관성이 우세한 영상에서 사용되는 예측 구조들을 제안하였다.

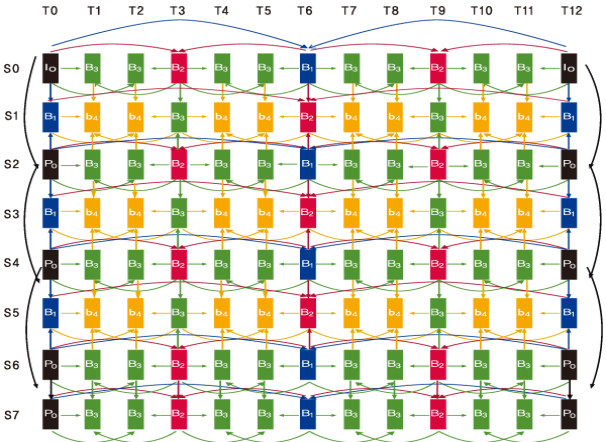
본 논문에서 제안하는 예측 구조는 다시점 비디오의 부호화 효율을 높이기 위하여 부호화되는 현재 화면과 현재 화면이 참조하는 참조 화면들과의 평균 거리와 B계층의 최대 인덱스와 각 B계층의 화면 수를 고려하였다.

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

2. 참조 예측 구조와기존방법들

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. 그러므로 다시점 비디오 부호화는 영상들 사이에 존재하는 시간적 상관성과 화면간 상관성을 이용하여 시간적으로 중복된 데이터와 시점간 중복된 데이터를 제거한다. 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 참조 예측 구조로 기술한다. 일반 사용자에게 보다 편리한 사용자 인터페이스 환경을 제공하기 위해서는 현재의 윈도우즈의 기반 사용자 인터페이스의 차원을 넘어서 사용자의 작업을 대행해 줄 수 있는 에이전트 시스템이 제공되어야 한다. 또한 에이전트 시스템서비스 확장과 사용보급을 위하여 응용을 위한 미들웨어 플랫폼에 대한 연구개발이 이루어져야 한다.



(그림 1) Fraunhofer-HHI의 계층적 B화면 구조

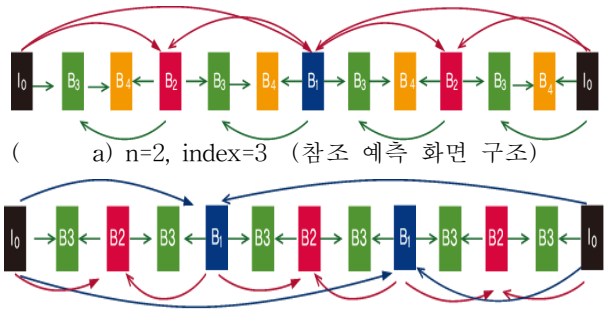
그림 1에서 Sn은 n번째 시점의 카메라를 의미하고 Tn은 시간적으로 n번째 화면을 나타낸다. 화살표는 이웃하는 화면들 사이의 예측 참조 관계를 나타낸다. 기존 시스템과의 호환성을 유지하기 위하여 다른 시점과 상관없이 독립적으로 복원할 수 있는 시점을 I시점, 부호화가 완료된 하나의 시점만을 참조하여 예측 부호화하는 시점을 P 시점 그리고 인접해 있는 두 개의 시점을 참조하여 예측 부호화하는 시점을 B시점이라고 한다. 그림 1의 S_0 은 I 시점 그리고 S_2, S_4, S_6, S_7들은 P 시점, S_1, S_3, S_5들은 B시점이다. 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.

표준 예측 구조의 계산량을 줄이기 위하여 여러 가지 예측 구조들이 제안되었다[6-8]. 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 Bi 계층에 속하는 화면 수를 고려하였다. Bi에서 i를 인덱스라 정의한다. 일반적으로 B계층 최대 인덱스는 GOP를 몇 개의 그룹으로 분할하느냐에 따라 달라진다.



(a) n=2, index=3 (참조 예측 화면 구조)

(b) n=3, index=3 (제안 예측 구조)

(그림 2) GOP 분할 그룹 수와 B계층 인덱스

그림 2에서 n은 길이가 12인 GOP를 여러 개의 작은 그룹으로 분할할 때 생성되는 그룹의 수로 B1 화면 수에 1을 더한 것과 같다. 그리고 Bi에서 i를 인덱스로 정의하였을 때, index는 B계층 최대 인덱스이다. 길이가 12인 GOP가 만들 수 있는 B계층의 최대 인덱스는 아래와 같다.

$$\text{Layer\_max} = \lceil \log_n GOPlength \rceil \qquad \text{(식 1)}$$

식 (1)에서 $GOPlength$는 GOP의 길이를 나타내고, n은 길이가 12인 GOP의 분할 그룹의 수를 나타낸다. 즉, n은 하나의 GOP를 몇 개의 그룹으로 나눌 것인가를 나타낸다. 그림 2 (a)는 길이가 12인 GOP를 2개 그룹으로 나누었을 때 생성되는 Layer_max는 4이므로 B계층 최대 인

덱스는 1에서 4사이의 정수이고, 그림 2(b)에서 GOP를 3개 그룹으로 나누었을 때 생성되는 Layer_max는 3이므로 B계층 최대 인덱스는 1에서 3사이의 정수이다.

<표 1> 분할 그룹 수와 각 B_i 계층의 화면 수와 평균 거리

	GOP 12 분할 그룹 수	
	그림 2(a)	그림 2(c)
	n=2	n=3
B1 계층	1개	2개
B2 계층	2개	3개
B3 계층	8개	6개
B4 계층		
평균거리	4.36	4.36

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 발생되는 비트량은 많아지고 영상의 화질이 좋다. 반면에 B계층 최대 인덱스가 클수록 발생되는 비트량은 적어지며 영상 화질도 좋지 않다. 그러므로 영상 화질과 발생되는 비트량을 고려하여 각 GOP에서 B계층의 최대 인덱스와 각 B계층의 화면 수를 결정하여야 한다.

영상 화질과 발생되는 비트량을 고려하여 각 GOP에서 B계층의 최대 인덱스와 각 B계층의 화면 수를 결정하기 위하여 부호화되는 현재화면과 현재화면이 참조하는 참조화면들과의 평균거리를 이용한다. 분할되는 그룹 수, B계층 최대 인덱스 그리고 B_i 계층의 화면 수에 따라 GOP에서 화면들을 부호화 할 때 부호화되는 현재화면과 현재화면이 참조하는 참조화면과의 거리가 달라진다. 부호화되는 현재화면과 현재화면이 참조하는 참조화면들과의 거리를 구하기 위하여 식 (2)를 이용한다.

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

$$Distance_S = \sum_{i=1}^{max} i_{Nmber} \times distance_i \quad \text{(식 3)}$$

$$Distance_{mean} = \frac{Distance_S}{GOPlength - 1} \quad \text{(식 4)}$$

식 (2)에서 N은 분할되는 그룹 수이고, GOPlength는 GOP 길이이며 i는 B계층 인덱스이다. 식 (2)에서 B_i 계층의 참조화면들과의 거리인 $distance_i$가 2미만이면 부호화되는 현재화면과 현재화면이 참조하는 참조화면들과의 거리를 2로 한다. B_i 계층의 $distance_i$가 2미만인 경우에도 부호화되는 현재화면이 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 현재화면과 현재화면이 참조하는 참조화면들과의 평균 거리가 길어진다. 식 (3)에서 max는 GOP를 분할하였을 때 생성되는 B계층 최대 인덱스를 의미한다. 예를 들면, 그림 2(a)에서 max는 3이다. 그리고 식 (3)에서 i_{Nmber} 는 B_i 계층의 화면 수를 의미하고 $distance_i$는 현재화면과 현재화면이 참조하는 참조화면들과의 거리이다.

그림 2(a)는 GOP를 두 그룹으로 분할한 구조로 분할되는 그룹 수 N=2이고 max=3이며 생성되는 B_1 계층의 화면 수는 1, B_2 계층의 화면 수는 2, B_3 계층의 화면 수는 8이다. 그리고 $distance_1$=12, $distance_2$ =6, $distance_3$=3이다. 그림 2(b)는 N=3이고 max는 3이며 생성되는 B_1 계층의 화면 수는 2, B_2 계층의 화면 수는 3, B_3 계층의 화면 수는 6이다. 식 (3)을 이용하여 각 예측 구조의 현재화면과 현재화면이 참조하는 참조화면들과의 거리의 합을 구하면 그림 2(a)의 경우 $Distance_S$는 1*12 + 2*6 + 8*3 이므로 48이다. 식 (4)을 이용하여 각 예측 구조의 현재화면과 참조화면들과의 평균거리를 구한다. 그림 2(a)의 경우 $Distance_{mean}$는 $\frac{48}{12-1}$ 이므로 4.36이고 그림 2(b)의 $Distance_{mean}$는 4.36이다.

본 논문에서 다시점 비디오의 부호화 효율을 높이기 위하여 부호화하는 현재화면과 현재 화면이 참조하는 참조화면들과의 평균 거리뿐만 아니라 B계층 인덱스와 그리고 각 B_i 계층에 속하는 화면 수를 고려하였다. 영상 화질과 발생되는 비트량을 고려하여 각 GOP에서 B계층의 최대 인덱스와 각 B계층의 화면 수를 결정하기 위하여 그림 2(a)의 참조 예측 구조의 최대 인덱스와 각 B계층의 화면 수 그리고 평균 거리를 기준으로 하였다. B계층의 최대 인덱스와 각 B계층의 화면 수 그리고 평균 거리는 영상 화질과 발생되는 비트량에 영향을 주므로 최대 인덱스는 참조 예측 구조와 같으면서 B계층의 최대 인덱스의 화면 수가 참조 예측 구조의 최대 인덱스의 화면수보다 적은 구조를 제안하였다. 본 논문에서 제안하는 예측 구조인 그림 2(b)의 예측 구조를 참조 예측 구조인 그림 2(a)와 비교하였을 때, B_i 계층의 화면 수는 다르지만 평균 거리와 B계층 최대 인덱스는 같다. 제안하는 예측 구조의 B_1 계층의 화면 수와 B_2 계층의 화면 수가 참조 예측 구조의 B_1 계층, B_2 계층의 화면 수 보다 많으므로 제안하는 예측 구조가 참조 에측 구조보다 영상 화질 면에서 좋고 비슷한 비트량을 보일 것이다.

4. 실험결과

제안하는 예측 구조의 성능을 확인하기 위하여 제안하는 예측 구조를 JMVM에서 구현하였다. 실험 영상으로 Exit, Ballroom, Uli를 사용하였다. 실험조건 표 2에 나타내었다.

<표 2> 실험 조건

영상	영상의 크기	기본 QP	탐색범위	프레임 수
Exit	640*480	37	±96	100
Ballroom	640*480			
Uli	1024*768			

<표 3> Exit 실험 결과

	참조 예측 구조		제안 예측 구조	
카메라 번호	PSNR (db)	bit rate (Kbps)	PSNR (db)	bit rate (Kbps)
0	35.1402	113.536	35.1984	113.058
1	34.0961	60.308	34.1864	59.076
2	34.8471	95.966	34.9181	95.8
3	33.7928	72.814	33.864	70.764
4	34.2998	118.052	34.3741	118.05
5	33.6858	89.168	33.7472	87.45
6	33.7063	149.758	33.7872	149.19
7	33.6582	131.69	33.7121	129.834

<표 4> Uli 실험 결과

	참조 예측 구조		제안 예측 구조	
카메라 번호	PSNR (db)	bit rate (Kbps)	PSNR (db)	bit rate (Kbps)
0	31.5464	942.772	31.6826	934.506
1	29.1249	753.902	29.2822	744.316
2	32.0423	798.246	32.1832	792.51
3	30.22	592.552	30.3479	581.81
4	32.95	609.114	33.0988	603.422
5	31.6668	417.384	31.8147	409.764
6	33.1867	552.848	33.3269	547.406
7	32.2128	675.804	32.319	674.674

<표 5> Ballroom 실험 결과

	참조 예측 구조		제안 예측 구조	
카메라 번호	PSNR (db)	bit rate (Kbps)	PSNR (db)	bit rate (Kbps)
0	32.1016	295.704	32.1929	298.32
1	31.2187	156.412	31.3004	158.25
2	32.157	245.95	32.267	250.278
3	31.2049	152.912	31.2941	154.648
4	31.5052	268.79	31.6076	273.244
5	31.7143	172.128	31.7869	172.64
6	31.8597	265.312	31.966	270.094
7	31.1416	245.28	31.239	249.946

본 논문에서 제안한 예측 구조의 성능과 참조 예측 구조의 성능을 표3, 표4, 표5에 제시하였다. 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 참조 예측 구조보다 영상 화질 면에서 평균 0.07(dB) 성능 향상을 보였으며 발생 비트량에 있어서 평균 1.008(Kbps)의 감소를 보였다.

표 4에 제시된 Uli 영상의 실험 결과를 살펴보면 제안한 예측 구조가참조 보다 화질 면에서 평균 0.13(dB) 성능 향상을 보였으며 발생 비트량에 있어서 평균 6.776(Kbps) 감소를 보였다. 표 5에 제시된 Ballroom 영상의 실험 결과에서 제안한 예측 구조는 영상 화질 면에 있어서 참조 예측 구조보다 평균 0.10(dB) 성능 향상을 보였지만 발생 비트량에 있어서 평균 3.116(Kbps) 증가를 보였다.

5. 서론

본 논문에서는 다시점 비디오의 부호화 효율을 높이기 위하여, 즉 영상 화질을 향상시키면서 발생되는 비트량을 줄이기 위한 예측 구조를 제안하였다. 제안한 예측 구조는 부호화되는 현재 화면과 현재 화면이 참조하는 참조 화면들과의 평균거리뿐만 아니라 B계층 최대 인덱스 그리고 각 B_i계층의 화면 수를 고려하였다. 제안하는 예측 구조와 참조 예측 구조는 현재 화면과 현재 화면이 참조하는 참조 화면들과의 평균거리와 최대 인덱스는 같지만 제안하는 예측 구조의 B_1계층의 화면 수와 B_2계층의 화면 수가 참조 예측 구조의 B_1계층, B_2계층의 화면 수 보다 많으므로 영상 화질 면에서 제안하는 예측 구조가 참조 예측 구조보다 약 0.07~0.13 (dB) 성능 향상을 보였다. 그리고 초당 비트량에 있어서 약 +3 ~ - 6.5(Kbps) 감소하였다.

Acknowledgment

"이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2010-0024120)

참고문헌

[1] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. kauff, P. Eisert, and T. Wiegand, "3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards," in Proc. of IEEE International Conference on Multimedia and Exposition, Jul. 2006.

[2] ISO/IEC JTC1/SC29/WG11 N10357, "Vision on 3D Video", Feb. 2009

[3] 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," in Proc. of IEEE International Conference on Multimedia and Exposition, Jul. 2006.

[4] 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.

[5] ISO/IEC JTC1/SC29/WG11, "Joint Multiview Video Model (JMVM) 8,"Doc. N9762, May 2008

[6] P.-K. Park, K.-J. Oh, Y.-S. Ho. Efficient view-temporal prediction structures for multi-view video coding [J]. Electronics Letters, 2008, 44(2): 102-103.

[7] R.L. He, "A Multiview Video Coding Method with Adaptive Selection of Reference Frame Modes," Journal of Computer-aided Design And Computer Graphics..30 (12), pp. 2205-2211, Dec. 2007.

[8] F. Lu, P. An, Z. Zhang, L. Shen, "Multi-view Video Coding Based on Sequence Correlation", Audio Language and Image Processing (ICALIP), pp. 1227-1232 , 2010