

논문 2010-47IE-2-4

# H.264/AVC Scalability Extension의 부호화 효율 향상 기법

## ( A Method for Improvement of Coding Efficiency in Scalability Extension of H.264/AVC )

강 창 수\*

( Chang Soo Kang )

### 요 약

본 논문에서는 H.264/AVC SE(Scalability Extension)의 부호화 효율을 향상시킬 수 있는 방법을 제안하였다. 제안한 방법은 율-왜곡 최적화 기법(Rate-Distortion Optimization, RDO)이 적용된 JM(Joint Model)의 FME(Fast Motion Estimation)를 대상으로 예측된 움직임 벡터의 방향성을 고려하여 탐색영역을 결정한 후 적응적인 후보 나선형 탐색을 수행한다. 다양한 영상들을 대상으로 실험한 결과 기존의 예측 방식에 최대 80%의 연산량을 줄일 수 있음을 확인하였다. 이에 따른 화질 열화는 평균 0.05dB~0.19dB에 불과하며, 압축률은 평균 0.58%의 미미한 감소를 보임으로써, 제안한 방법이 매우 효율적임을 확인하였다.

### Abstract

This paper proposed an efficient algorithm to reduce the amount of calculation for Scalability Extension which takes a great deal of the operational time in H.264/AVC. This algorithm decides a search range according to the direction of predicted motion vector, and then performs an adaptive spiral search for the candidates with JM(Joint Model) FME(Fast Motion Estimation) which employs the rate-distortion optimization(RDO) method. Experimental results by applying the proposed method to various video sequences showed that the process time was decreased up to 80% comparing to the previous prediction methods. The degradation of video quality was only from 0.05dB to 0.19dB and the compression ratio decreased as small as 0.58% in average. Therefore, we are sure that the proposed method is an efficient method for the fast inter prediction.

**Keywords :** H.264/AVC, RDO, Video Quality, Fast Inter Prediction, JM

### I. 서 론

현재 멀티미디어 기술은 급속도로 발전하면서 다양한 코딩기술 및 처리, 분배 능력과 함께 콘텐츠 제작자와 소비자들에게 제공되고 있다. 동시에 유니쿼터스 통신환경 구조(Ubiquitous Communication Infrastructure)는 이러한 멀티미디어 정보들을 다양한 통신망과 단말에 대응하여 UMA(Universal Multimedia Access) 기반의 "Anytime, Anywhere, Anyone"의 서비스가 가능하도록 요구되어진다<sup>[1]</sup>. 이러한 다양한 수신자의 요구조건

에 따른 UMA 서비스를 충족시키기 위해서 ISO/IEC JTC1 SC29 WG11의 MPEG-21 Part 13 SVC(Scalable Video Coding)를 위한 연구가 진행되었으며<sup>[2]</sup>, 현재는 ISO/IEC MPEG과 ITU-T VCEG의 JVT(Joint Video Team)에 의해 H.264/AVC AMD1 Scalability Extension으로 전환되어 표준화 과정이 진행 중에 있으며 최근 Working Draft 3.0 of 14496-10/AMD1이 발표되었다<sup>[3]</sup>.

SVC(Scalable Video Coding)는 하나의 비디오 스트림으로 다양한 전송 네트워크와 수신 단말에 적응적 서비스가 가능하게 하기 위한 비디오 부호화 방법으로, 기존의 관련 국제 표준 및 활동으로는 MPEG-2 Scalable Profile<sup>[4]</sup>, MPEG-4 Scalable Profile/FGS<sup>[5]</sup>,

\* 정회원, 유한대학 전자정보과  
(Yuhan University Dept. Electronic & Information Engineering)  
접수일자: 2010년1월28일, 수정완료일: 2010년6월7일

MPEG-21 Part 13 SVC 등이 있다. MPEG-2 Scalable Profile은 기본계층(Base Layer)과 향상계층(Enhancement Layer)의 구조를 가지면서 공간적(Spatial), 시간적(Temporal), 화질(SNR)에 대한 스케일러빌리티(Scalability)를 제공하지만 동시에 한가지 스케일러빌리티만 제공이 가능하다는 단점을 가지고 있다. MPEG-4 Scalable Profile은 FGS(Fine Granularity Scalability)를 이용하여 화질에 대한 스케일러빌리티를 제공하고 있으며, MPEG-21 Part 13 SVC는 이러한 스케일러빌리티와 더불어 네트워크 상에서의 강건성, 기본계층 호환성, 저 복잡도 코덱, 단대단 지연, 화질 선택 및 코딩 효율을 고려한다. 현재 표준화 과정이 진행 중인 H.264/AVC AMD1 Scalability Extension은 다양한 비트스트림을 생성하기 위해 동영상 압축 표준인 H.264/AVC<sup>[6]</sup>를 기반으로 한 계층적 접근 방식을 사용함으로써 스케일러빌리티를 제공한다. 이 중 시간적 스케일러빌리티를 위해 MCTF(Motion Compensated Temporal Filtering)와 Hierarchical B pictures를 사용하여 고정된 GOP 구조 내에서 영상들 간의 시간축으로의 재분배 과정을 수행하여 기본계층과 향상계층을 생성한다. 최근 H.264/AVC SE의 부호화 효율 향상을 위해 국내연구기관인 ETRI에 의해 영상간의 특성을 고려하여 가변적인 GOP 구조에 대해 MCTF를 수행하는 Adaptive GOP structure for SVC가 제안되었다<sup>[7]</sup>.

본 논문에서는 H.264/AVC SE의 부호화 효율을 개선시키기 위해 고정된 GOP 내에서 영상들 간의 특성을 고려하여 최적의 sub-GOPs 구조로 재구성하여 Hierarchical B pictures를 수행하는 AGS-HBP(Adaptive GOP Structure Hierarchical B Pictures) 알고리즘을 제안하였다. 본 논문은 II장에서 H.264/AVC SE 및 MCTF와 Hierarchical B pictures의 동작을 기술하고, III장에서는 AGS-HBP 알고리즘을 제안한다. 그리고 IV장에서는 제안한 알고리즘의 실험 결과를 분석하고, V장에서 결론을 맺는다.

## II. H.264/AVC SE 및 MCTF와 Hierarchical B pictures

### 2-1 H.264/AVC SE(Scalable Extension)

H.264/AVC SE의 공간적, 시간적, 화질에 대한 스케일러빌리티를 제공하기 위한 부호화기의 구조를 그림 1에 나타내었다<sup>[3]</sup>. H.264/AVC SE의 부호화기는

H.264/AVC Codec을 기반으로 SVC를 접목시키기 위해 추가적인 기본계층과 향상계층의 계층적 접근 방식을 통하여 다양한 비트스트림을 생성 및 추출한다. 공간적 스케일러빌리티를 위해 Spatial Decimation/Interpolation 시간적 스케일러빌리티를 위해 MCTF/Hierarchical B pictures, 화질 스케일러빌리티를 위해 CGS(Coarse Grain Scalability)/FGS(Fine Granular Scalability)에 의해 수행되어진다.

입력 영상의 공간적 다운샘플링(Spatial Decimation)을 통한 공간적 기본계층과 향상계층들에 대해서 각각 독립적으로 화면간 예측(Inter Prediction) 방법과 함께 MCTF/Hierarchical B pictures에 의한 시간축으로의 영상 분해 과정 및 화면내 예측(Intra Prediction)을 수행한다. 낮은 공간적 계층으로부터의 움직임 예측 정보는 높은 공간적 계층의 움직임 예측을 위해 사용되어진다(Inter-Layer Motion Prediction). 공간적 계층들 사이에 존재하는 텍스처(Texture) 정보들의 중복성을 제거하기 위해 기본 계층으로부터 향상계층으로의 공간적 업샘플링(Spatial Interpolation)을 통해 텍스처 정보들의 차이 값을 구한 후(Inter-Layer Intra Texture/Residual Prediction), 향상계층들은 이러한 차이 값들을 이용하여 정수 4x4 DCT(Discrete Cosine Transform), 양자화(Quantization) 그리고 엔트로피 부호화 과정인 CAVLC(Context Adaptive Variable Length Coding)/CABAC(Context-based Adaptive Binary Arithmetic Coding)을 통하여 부호화된다. 그리고 CGS/FGS에 의해 화질 향상계층은 기본계층과 향상계층들 사이의 양자화 차이 값들로 이루어져 부호화된다. 추출기(Extractor)는 수신단의 전송 통신망과 단말의 상태에 따라 공간적, 시간적, 화질의 기본계층과 향상계층들의 적절한 조합을 통하여 다양한 비트스트림을 생성 및 추출함으로써 스케일러빌리티를 제공하게 된다.

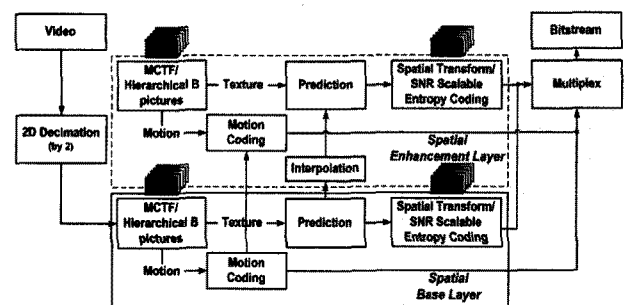


그림 1. H.264/AVC SE 부호화기의 구조  
Fig. 1. Encoder Structure of H.264/AVC SE.

2-2 MCTF

(Motion Compensated Temporal Filtering)

MCTF(Motion Compensated Temporal Filtering)는 Ohm에 의해 처음으로 제안된 알고리즘으로<sup>[8]</sup>, 영상들을 시간축으로의 중복성을 제거하기 위한 기술로 그림 2와 같은 피라미드형 분해 구조(Pyramidal decomposition structure)를 갖는다.

그림 2와 같이, H.264/AVC SE는 고정된 GOP 단위로 Haar와 5/3필터(filter)의 리프팅 기반(Lifting-based) MCTF를 적용하고 있으며, 각 레벨(Level)에 대해 예측(Prediction) 및 갱신(Update) 과정을 수행하여 각각 고주파 영상(High-pass Frame)과 저주파 영상(Low-pass Frame)을 생성한다. 예측 과정은 현재 영상을 기준으로 이전과 이후의 영상을 참조하여 움직임 예측(Motion Estimation)/보상(Motion Compensation) 과정을 통해 차영상 즉 고주파 영상을 만들어낸다. 갱신 과정은 이전의 예측 과정에서 구한 움직임 벡터(Motion Vector)를 이용하여 고주파 영상을 참조하여 움직임 보상을 통해 원 영상의 고주파 성분이 제거된 저주파 영상을 생성한다. 이와 같은 예측 및 갱신 과정은 각 레벨에 대해 수행되어지며 GOP = 2N일 경우, N 레벨까지 수행하고 마지막 레벨의 저주파 영상과 각 레벨의 고주파 영상이 부호화기의 다음 단계로 이동한다. 최종적인 저주파 영상은 다음 GOP의 MCTF를 수행할 시 예측 과정을 위해 사용되며, 복호화 시 이러한 영상들의 선택적인 추출로부터 시간적 스케일러빌리티가 제공된다.

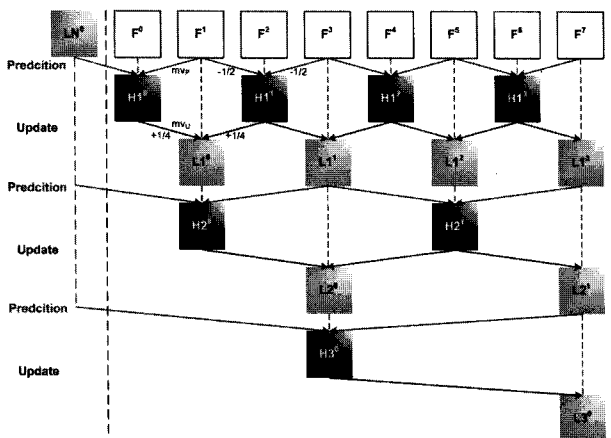


그림 2. H.264/AVC SE의 MCTF 구조  
Fig. 2. MCTF structure of H.264/AVC SE.

2-3 Hierarchical B pictures

Hierarchical B pictures는 MCTF와 유사하지만 갱신

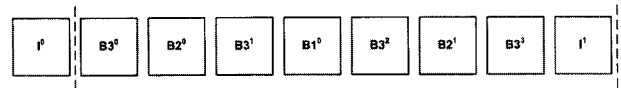


그림 3. H.264/AVC SE의 Hierarchical B pictures 구조  
Fig. 3. Hierarchical B pictures structure of H.264/AVC SE.

과정이 없이 예측 과정을 수용하면서 시간적 스케일러빌리티를 제공한다. 이 알고리즘을 사용한 H.264/AVC SE 부호화기의 비트스트림은 기존의 H.264 /AVC Main profile 복호화기와 호환 가능하며, 그림 3에 알고리즘 수행 방법을 보이고 있다. 첫 번째 영상은 IDR(Instantaneous Decoding Refresh) picture로 부호화되며, 이후의 영상들은 GOP 내에서 “B...BP” 혹은 B...BI“의 비디오 구조로 부호화된다. 첫 번째 레벨의 B1 영상은 주변 I 영상으로부터 예측되어지며, 다음 레벨의 Bi 영상들은 이전 레벨의 I 와 Bj ( $j < i$ )로부터 양방향 예측되어진다. 이들 영상은 RPLR(Reference Picture List Reordering)과 MMCO(Memory Management Control Operation) 제어 명령어와 함께 부호화되어 전송된다. 이와 같이 Hierarchical B pictures를 수행한 후 마지막 레벨의 결과인 I, B, P 영상들은 MCTF와 마찬가지로 부호화기의 다음 단계로 이동한다.

III. 제안한 AGS-HBP(Adaptive GOP Structure-HBP) 알고리즘

H.264/AVC SE는 고정된 GOP 구조 내에서 Hierarchical B pictures를 수행하여 시간적 스케일러빌리티를 제공하는데, 움직임이 많은 영상들을 부호화시 현재 영상과 참조 영상과의 중복성이 감소하여 차영상(B frames)내 인트라 처리가 많아진다. 이는 전체적인 코덱의 부호화 효율을 감소시킨다. 이러한 문제점을 해결하기 위해 본 논문에서는 고정된 GOP 구조 내에서 영상들 간의 특성을 고려하여 최적의 sub-GOPs로 재구성하여 Hierarchical B pictures를 수행함으로써 H.264/AVC SE의 부호화 효율을 향상시킬 수 있는 알고리즘을 제안하였다.

3-1 AGS-HBP의 sub-GOP 단위 생성 모델

그림 4는 8 프레임으로 고정된 GOP 내에서 sub-GOP 단위 생성 모델에 의한 가능한 sub-GOPs 구조를 나타내었다.

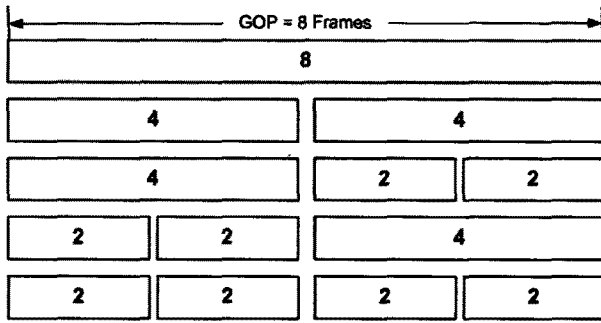


그림 4. sub-GOP 단위에 의한 sub-GOPs 재구성 구조  
Fig. 4. sub-GOPs reconstructed structure with sub-GOP unit.

$2^N=8$  ( $N:3$ ) 프레임의 GOP 구조는  $N=N-1$  ( $0 \leq 1 < N$ )의 sub-GOP 단위들을 생성할 수 있으며, 상위 sub-GOP는 하위 sub-GOP의 x2 (scaling factor) 연산을 통하여 같은 프레임율을 갖도록 구성된다. 이러한 sub-GOP 단위는 현재의 GOP를 발생 가능한 경우의 조합을 통하여 sub-GOPs로 재구성한다.

3-2 AGS-HBP 알고리즘

H.264/AVC SE의 부호화 효율을 개선시키기 위해, 고정된 GOP 구조 내에서 영상들 간의 특성을 고려하여 최적의 sub-GOPs로 재구성하여 Hierarchical B pictures를 수행하는 AGS-HBP 알고리즘은 그림 5와 같다.

AGS-HBP 알고리즘은 먼저 고정된 GOP 구조를 갖는 원래의 비디오 영상들에 대해 위의 그림 4에서 제안한 sub-GOP 단위 생성 모델을 이용하여 가능한

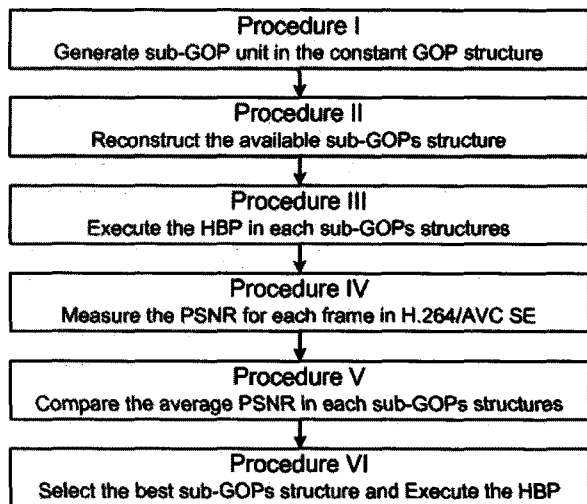


그림 5. AGS-HBP 알고리즘  
Fig. 5. AGS-HBP Algorithm.

```

Procedure (Seek the best sub-GOPs structure)
// sub-GOP_N = Average PNSR in N frames
begin
gop mode = min(sub-GOP_8, (sub-GOP_4_1 +
sub-GOP_4_2)/2, (sub-GOP_2_1 + sub-GOP_2_2 +
sub-GOP_2_3 + sub-GOP_2_4)/4)

If(gop mode == sub-GOP_8) then
  (Select 8 frame's sub_GOPs structure)
else
  gop mode_1 = min(sub-GOP_4_1, (sub-GOP_2_1 +
sub-GOP_2_2)/2)
  gop mode_2 = min(sub-GOP_4_2, (sub-GOP_2_3 +
sub-GOP_2_4)/2)
  if(gop mode_1 == sub-GOP_4_1) then
    (Select 1st 4 frame's sub_GOPs structure)
  else
    (Select 1st, 2nd 2 frame's sub_GOPs structure)
  if(gop mode2 == sub-GOP_4_2) then
    (Select 2nd 4 frame's sub_GOPs structure)
  else
    (Select 3rd, 4th 2 frame's sub_GOPs structure)
}
    
```

그림 6. GOP=8인 AGS-HBP의 최적의 sub-GOPs 구조 선택 과정  
Fig. 6. Select Procedure for the best sub-GOPs structure using AGS-HBP with GOP=8.

sub-GOPs 구조를 재구성한다. 이러한 각각의 sub-GOPs 구조에 대하여 Hierarchical B pictures를 수행한 후, 비디오 원 영상과 H.264/AVC SE Codec에 의해 I, B, P 영상들의 부/복호화된 결과 영상들의 화질을 측정한다. 영상 화질을 측정하기 위한 화질 평가 도구로는 PSNR(Peak Signal to Noise Ratio)을 이용하며, MSE(Mean Squared error) 연산 과정을 위해 사용되는 영상 픽셀 값들은 비디오 원 영상과 sub-GOPs별 Hierarchical B pictures를 통해 얻은 I, B, P 영상들의 부/복호화 수행 후 복원된 영상 내 픽셀 값들이며, 이러한 픽셀들의 차이 값을 이용하여 각 영상들의 PSNR을 얻게 된다.

각 영상들의 PSNR은 그림 6에서 Pseudo Code로 나타난 것과 같이 각 sub-GOPs 구조의 평균 PSNR을 비교하여 최적의 sub-GOPs 구조를 선택한다.

이와 같은 방법에 의해 최적화된 sub-GOPs 구조는 Hierarchical B pictures를 수행하여 H.264/AVC SE의 부호화 효율을 개선시킬 수 있는 비트스트림을 생성해 낸다.

### 3-3 AGS-HBP 알고리즘에 의한 시간적 스케일러 빌리티

그림 7과 같이, 시간적 스케일러빌리티를 위해 8 프레임의 고정된 GOP 구조 내에서 최적의 sub-GOPs 구조가 4+2+2인 경우, 다음과 같은 방법에 의해 추출기로부터 선택되어 복호화기에 전송된다.

Full Resolution : I<sup>1</sup>, B1<sup>0</sup>, B2<sup>0</sup>, B2<sup>1</sup>, I<sup>2</sup>, B1<sup>1</sup>, I<sup>3</sup>, B1<sup>2</sup>

1/2 Resolution : I<sup>3</sup>, B1<sup>2</sup>, I<sup>2</sup>, B1<sup>1</sup>

1/4 Resolution : I<sup>3</sup>, B1<sup>2</sup>

1/8 Resolution : I<sup>3</sup>

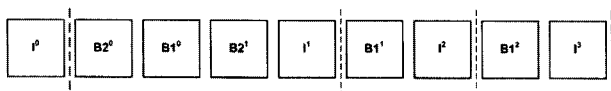


그림 7. AGS-HBP에 의해 재구성된 최적의 sub-GOPs(4+2+2) 구조  
 Fig. 7. Reconstructed the best sub-GOPs(4+2+2) structure using AGS-HBP.

### IV. 실험 결과

본 논문에서 제안한 알고리즘은 SVM 3.0 참조 소프트웨어<sup>[9]</sup>를 이용하여 MPEG에서 정의한 테스트 영상들을 대상으로 검증하였다. 테스트 영상들은 각각 기존의 고정된 GOP 구조 방식의 SVM과 제안한 AGS-HBP 알고리즘을 적용시킨 SVM을 이용하여 부/복호화 하였다. 부호화시 8 프레임의 고정된 GOP와 함께 기본 계층의 해상도는 QCIF와 향상 계층은 CIF인 조건에서 가변적인 비트율을 적용하였다.

그림 8, 9는 “Football”과 “Crew” 영상에 대해 이와 같은 방법을 사용하여 얻은 복호화된 영상들의 평균 P

SNR을 나타내었다. 움직임이 많은 영상들에 있어서 기존의 SVM에 비해 제안한 AGS-HBP 알고리즘의 SVM이 영상들의 평균 PSNR을 향상시킨다는 것을 확

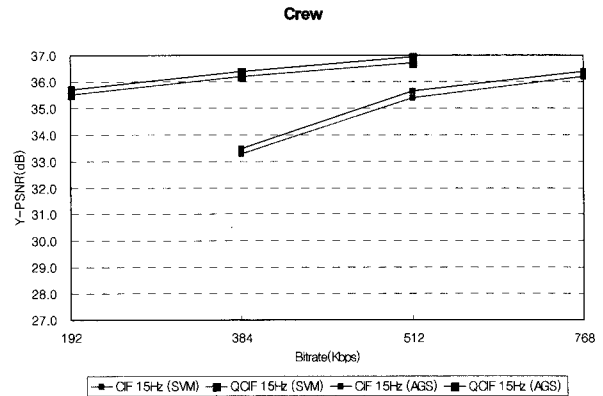


그림 9. Crew 영상에 대한 SVM과 AGS-HBP의 PSNR 비교  
 Fig. 9. SVM vs AGS-HBP PSNR compare for Crew image.

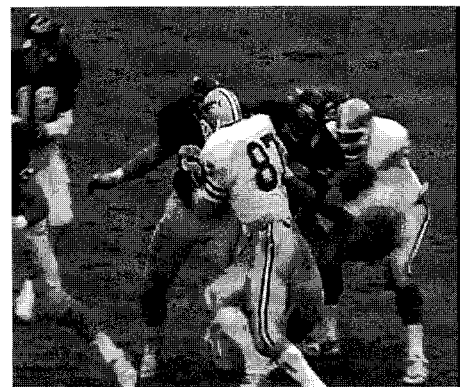


그림 10. Football 영상에 대한 AGS-HBP 기반의 SVM 적용 예  
 (a) SVM 복호화 영상  
 (b) AGS-HBP 복호화 영상

Fig. 10. Example of AGS-HBP based SVM for Football image.  
 (a) Decoded image in SVM  
 (b) Decoded image in AGS-HBP based SVM.

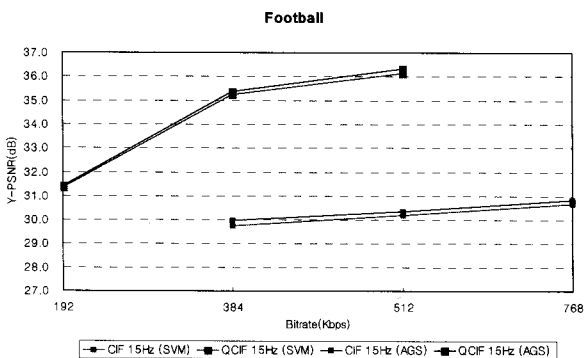


그림 8. Football 영상에 대한 SVM과 AGS-HBP의 PSNR 비교  
 Fig. 8. SVM vs AGS-HBP PSNR compare for Football image.

인할 수 있다. 실험 결과 0.05~0.25dB의 화질 향상을 검증하였으며, 이로 인해 H.264/AVC SE의 부호화 효율을 개선시킬 수 있다는 것을 입증하였다.

그림 10은 CIF 해상도를 갖는 원영상에 대해 512Kbps의 비트율의 환경 조건에서 기존의 SVM과 제안한 AGS-HBP의 SVM과의 복호화 된 결과를 보이고 있다. 이는 제안한 방법이 기존의 SVM에 비해 시각적으로 화질을 향상시킴을 알 수 있다.

## V. 결 론

본 논문에서는 AGS-HBP 알고리즘을 기반으로 H.264/AVC SE 부호화기의 효율을 향상시키는 방법을 제안하였다. 이를 위해 고정된 GOP 내에서 영상들 간의 특성을 고려하여 최적의 GOP 구조로 재구성하여 Hierarchical B pictures를 수행하였다. 제안한 알고리즘은 복호화기의 구조를 변경시키지 않으며, 실험을 통하여 기존의 SVM에 비해 약 0.05~0.25dB의 화질 개선을 확인할 수 있었다. 이러한 AGS-HBP 알고리즘을 기반으로 다양한 비디오 영상의 특성에 따라 적응적이며 유동적인 H.264/AVC SE 부호화기는 다양한 스케일러빌리티를 제공하며 동시에 부호화 효율을 향상시킬 수 있다. 따라서 본 논문에서 제안한 방법은 현재 국제 표준화가 진행 중인 H.264/AVC SE의 성능 향상을 위해 유용하게 사용될 수 있으며, 미래지향적인 비디오 코딩 분야에서 매우 중요한 요소 기술이 될 것으로 사료된다.

## 참 고 문 헌

- [1] J. Bormans, J. Gelissen, and A. Perkis, "MPEG-21: The 21st Century Multimedia Framework", IEEE Trans. on Signal Processing Magazine, vol. 20, pp. 53-62, 2003.
- [2] ISO/IEC TR 21000-1:2001(E), "Part 1: Vision, Technologies and Strategy", 2000.
- [3] J. Reichel, H. Schwarz, and M. Wien, "Working Draft 3 of ISO/IEC 14496-10:2005/AMD1 Scalable Video Coding", ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6 Document N7310, 2005.
- [4] ISO/IEC 13818-2, "Generic Coding of moving pictures and associated audio information-Part 2: Video", also ITU-T Recommendation H.262, 1995.
- [5] ISO/IEC 14496-2, "Coding of audiovisual objects-Part 2: Visual", 1998.
- [6] ISO/IEC 14496-10 "Coding of Audiovisual Objects-Part 10: Advance Video Coding", 2003, also ITU-T Recommendation H.264 "Advanced Video Coding for Generic Audiovisual Services", 2003.
- [7] G. H. Park, M. W. Park, S. Y. Jeong, J. H. Cha, K. H. Kim, and K. W. Hong, "Adaptive GOP structure for SVC", ISO/IEC JTC1/SC29/WG11 Document M11563, 2005.
- [8] J. R. Ohm, "Three Dimensional Subband Coding with Motion Compensation", IEEE Trans. on Image Processing, vol. 3, pp. 559-571, 1994.
- [9] "SVM 3.0 Software", ISO/IEC JTC1/SC29/WG11 Document N6717, 2004.

## 저 자 소 개

강 창 수(정회원)  
대한전자공학회 논문지  
2010년 3월 제47권 IE편 제1호  
(논문 2010-47IE-1-5)