

특집논문-06-11-2-01

# 다중 블록 크기의 움직임 예측과 SPECK을 이용한 고정 화질 움직임 보상 시간영역 필터링 동영상 압축

박 상 주<sup>a)†</sup>

## Constant Quality Motion Compensated Temporal Filtering Video Compression using Multi-block size Motion Estimation and SPECK

Sang Ju Park<sup>a)†</sup>

### 요 약

움직임 보상을 적용한 시간 영역 필터링(MCTF)을 이용한 화질 보장형의 새로운 동영상 압축 방식을 제안한다. SPECK은 그 자체의 단순한 알고리즘으로 인하여 빠른 동작 속도를 가지면서도 동시에 고주파 성분이 많은 영상의 압축에 탁월한 성능을 보여주는 우수한 웨이블릿 변환 기반의 영상 압축기법이다. 또한 제안한 계층적 구조의 다중 크기 블록 움직임 예측은 비교적 낮은 연산량에도 불구하고 기존의 고정 블록 크기의 움직임 예측기보다 우수한 성능을 보인다. 본 논문에서는 이러한 낮은 복잡도의 기술을 MCTF 기반 동영상 압축에 적용하여, 다중 재생률까지 지원이 가능한 동영상 압축 방식을 구현하였으며, H.263 압축방식에 비해 우수한 압축 성능을 보임을 확인하였다.

### Abstract

We propose a new video compression method based on MCTF(motion compensated temporal filtering) with constant quality. SPECK is an efficient image compression coding method of encoding DWT coefficients. Especially SPECK method is very efficient for coding the motion compensated residual image which usually has larger amounts of high frequency components than the natural images. And proposed multi block size hierarchical motion estimation technique is more efficient than classical block matching algorithm with fixed block size both in estimation precision and operation costs. Proposed video method based on MCTF video compression can also support multi-frame rate decoding with reasonable complexity. Simulation results showed that proposed method outperforms H.263 video compression standard.

Key Words : MCTF(Motion Compensated Temporal Filtering), 웨이블릿 변환, SPECK, 다중블록크기 움직임 예측, 동영상압축

### I. 서 론

다양한 멀티미디어 정보의 보급에 따라 효율적인 동영상

압축 기법에 대한 요구는 계속 증대되어 왔다. 이러한 요구에 발맞추어 MPEG<sup>[1]</sup> 및 MJPEG<sup>[2]</sup>과 같은 동영상 및 음향 압축의 국제 표준이 제정되었고, 이러한 국제 표준을 사용하는 HDTV 및 디지털 위성 방송과 같은 고품질 동영상 서비스가 제공되고 있다. 이러한 고품질 동영상 서비스와 더불어 PDA나 이동통신 단말기와 같은 저해상도, 저전송

a) 홍익대학교, 전자전기공학부

School of Electronics and Electrical Engineering, Hongik University

† 교신저자 : 박상주(sjpark@hongik.ac.kr)

※ 본 논문은 서울시 전략사업 혁신클러스터 육성 지원사업에 의하여 지원되었음

를 통신 기기를 사용한 동영상 서비스에 대한 요구도 증대되고 있다. 이러한 저해상도 응용을 위하여 H.263<sup>[3]</sup>과 같은 표준도 마련되었고, CDMA2000 이나 IMT2000 등의 무선 이동통신 서비스의 보급과 더불어 저 전송률 동영상 서비스에 대한 관심도 크게 증가하고 있다. 일반적인 현대의 영상 압축은 시간상의 중복을 제거하기 위해 움직임 예측과 보상을 이용하는 예측 부호화 방식과, 공간상의 중복을 제거하기 위해 변환기와 양자화기를 이용하는 변환 부호화 방식이 결합된 Hybrid 형식을 취한다<sup>[4]</sup>. MPEG, H.263 과 같이 현재 널리 사용되는 동영상 압축 표준들이 이러한 Hybrid 구조로 동작하며, 변환기법으로는 2차원 DCT<sup>[5]</sup>가 사용된다. 이론적으로 변환계수의 에너지 집중 특성이 가장 우수한, 최적의 변환 기법은 K-L 변환<sup>[6]</sup>이지만, K-L 변환은 고속 변환 알고리즘이 존재하지 않으며 입력 신호에 따라 기저가 바뀌므로 실용적인 사용에는 부적합하다. 이에 비해 DCT 의 경우 고속 변환 알고리즘이 존재하면서 일반적인 자연 영상에 경우 K-L 변환과 거의 유사한 성능을 보여준다. 하지만 DCT 기반 영상 압축 기법은 HDTV 나 디지털 위성방송과 같은 고 전송률 고품질 시스템에는 적합하나, 저 전송률 시스템에서의 응용에는 적합하지 않다. 이것은 DCT 변환이 블록 기반 변환 기법이고, 블록변환 기반 압축에서는 저 전송률에서 블록화 현상이 발생하여 시각적으로 나쁜 화질을 보이기 때문이다. 이에 비해 웨이블릿 변환은 블록기반 영상 압축 기법에서 나타나는 블록화 현상이 없어 주관적 화질에서 유리하면서 객관적인 화질 평가 기준인 PSNR 관점에서도 우수한 성능을 보여준다. 이렇게 여러 우수한 특성을 가지는 웨이블릿 변환을 동영상에서도 이용하기 위한 연구가 다각도로 있어왔으며, 기본적으로 DCT에 비해 우수한 에너지 압축 성능으로 인한 압축률 향상을 이룬 SAMCoW<sup>[7]</sup> 기법 등이 있었다. 하지만 일반적인 Hybrid 방식의 동영상 압축 방식의 경우 시간 영역 중복제거를 위해 사용하는 움직임 예측기가 기본적으로 순환적(close-loop) 구조를 이룸으로 인한 에러에 약한 성질을 지니고 있다. 전송 에러에 대해 강인한 내성은 네트워크를 통한 동영상 서비스에서 주요한 성질이며 본 논문에서는 때문에 에러에 강인한 특성을 가지는 웨이블릿 기반 동영상 압축 기법에 대해 연구 하였다. 본 논문에서

제안하는 방법은 시간 영역 중복제거에 움직임 보상을 적용한 웨이블릿 변환을 사용하여 기존에 순환적 구조를 제거해 에러에 강인하면서 웨이블릿 변환이 제공하는 다중 해상도를 지원함은 물론 다중 재생률(multi frame rate)까지 지원하는 특성을 갖는다.

움직임 보상을 위해서는 기존에 최저화질 보장형 동영상 부호화를 위하여 제안하였던 계층적 구조의 다중 크기의 움직임 블록을 이용한 움직임 예측을 이용하며<sup>[8]</sup>, 웨이블릿 계수를 부호화 하는데 SPECK<sup>[9]</sup>을 이용하였다. 제안한 방식은 상대적으로 낮은 연산량으로 우수한 압축 성능과 전송 에러에 강인함 그리고 다중 재생률까지 지원하는 특성을 갖고 있다.

본 논문은 2 장에서 SPECK기법 3장에서 MCTF(Motion Compensated Temporal Filtering)의 개요에 관하여 간단히 살펴보고, 4 장에서 제안하는 MBMA(Multi-block size Matching Algorithm)을 설명한 뒤, 5 장에서 모의실험 결과를 제시하고 6 장에서 결론은 맺는다.

## II. SPECK (Set Partitioning in Embedded block)

웨이블릿 기반 영상 압축 기법의 성능은 사용되어지는 웨이블릿 필터는 물론 웨이블릿 변환 계수의 양자화 방법에 의해 크게 좌우되며, 과거에는 주로 단순한 스칼라 양자화기나 벡터 양자화기를 사용하는 기법들이 검토되었다. 그러나 1993년 J. Shapiro에 의해 제안된 제로트리(Zerotree) 기반 EZW(Embedded Zeortree Wavelet) 기법이 압축 성능이 매우 우수하며 PSNR scalability에 잘 적용된다는 것이 알려진 이후<sup>[10]</sup>, 웨이블릿 계수의 부호화 방법으로는 EZW와 SPIHT<sup>[11]</sup>등 여러 가지 변형된 형태의 제로트리 기반 부호화 기법들이 사용되고 있다.

옥타브(octave) 웨이블릿 변환된 영상은 그림 1과 같이 제로트리로 표시되며, 변환 계수 및 대역들 사이에는 그림에 보인 것과 같은 부모-자손 관계가 성립한다. 자연 영상의 경우 부모 계수가 작은 값을 가질 경우 그것의 후손 계수들도 작은 값을 가질 확률이 높다는 제로트리 가설이 잘 적용되며, 이와 같은 가설에 기초한 EZW 기법은 우수한 압축 성

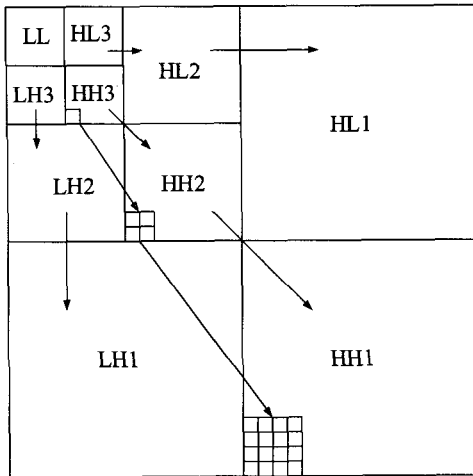


그림 1. 제로트리의 구조  
Fig. 1. Zerotree Structure

능과 아울러 주어진 비트율에서 최적의 압축 비트열을 얻을 수 있는 PSNR Scalability 특성으로 인하여 정지 영상 및 동영상 압축 기법의 기본으로 매우 광범위하게 적용되어져 왔다. 또한 기본적으로 EZW 에 제로트리 가설을 수용하면서 리스트 구조를 이용한 효율적인 동작으로 보다 우수한 압축 성능을 보이는 SPIHT 와 같은 기법들도 제안되었다. 이러한 웨이블릿 변환을 이용한 영상 압축 방법의 효율성은 일반적으로 제로트리 가설에 의지한다. 하지만 최근 연구들은 대역별 부모-자손 관계에서의 상관성이 SPIHT 나 EZW 의 높은 압축 성능에 결정적인 요소는 아니라는 것을 증명한다. 특히 동영상의 예측 오차영상의 경우 그림 2에서

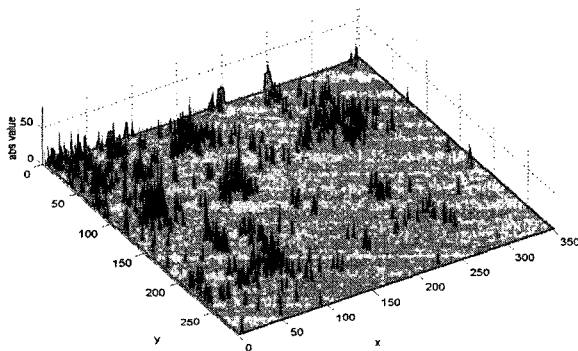


그림 2. 오차영상의 웨이블릿 변환 계수의 예  
Fig. 2. Example of wavelet transform coefficients of a predicted error frame

보듯이 웨이블릿 변환 시 저주파 대역으로의 에너지 집중이 매우 낮으며 부모 자손간의 상관성도 낮다.

또한 대역별 상관관계를 이용하는 부호화 기법의 경우에는 그림 3에 보인바와 같이 웨이블릿 변환 계수의 양자화 오차가 차츰 줄어드는 형태로 비트열의 embedding이 이루어져, 각 대역의 계수들이 서로 뒤섞인 형태로 비트열이 구성되어, PSNR scalability의 관점에서는 우수하나, 다중 해상도 지원(multi resolution scalability)의 관점에서는 적합하지 않은 구조를 갖고 있다.

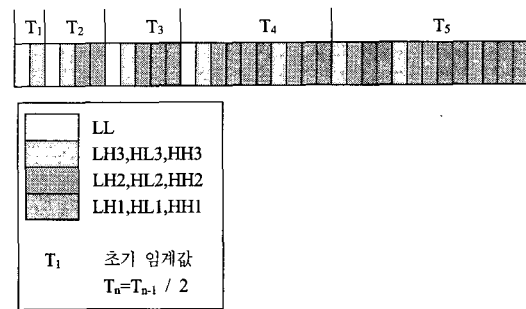


그림 3. EZW 부호화의 경우 비트열의 해상도별 대역 분포의 예  
Fig. 3. Example of EZW encoded bit stream distribution along resolution bands

이에 비해 SPECK<sup>[9]</sup>은 압축 성능이 기존의 EZW나 SPIHT에 비해 우수하며 다중해상도 지원에 유리한 영상 압축 방식으로 특히 동영상의 예측 오차 영상처럼 고주파 성분이 많은 영상을 부호화 할 때 우수한 압축 성능을 보여 준다. SPECK은 EZW 나 SPIHT와 마찬가지로 MSB에서 LSB로 개별 비트평면에 대해서 적용되며, 이는 이진 영상에서 중요 계수의 위치를 알려주는 방식으로 동작하는 양자화기 이다. SPECK의 기본적인 동작 방식은 웨이블릿 대역별로 독립적인 쿼드트리 부호화를 적용하는 것이다.

그림 4(a)의 예처럼 4x4 크기의 공간에 2개의 중요계수가 존재할 때 이를 쿼드트리로 표현 하면 그림 4(b)가 된다. 이 쿼드 트리를 비트열로 표현하기 위해서 검은색 노드는 1로 하얀색 노드는 0으로 표시 하며 루트 노드에서 시작해서 좌에서 우로 검사하여, 검은색 노드를 만날 때는 하위가지로 내려간다. 따라서 그림 4에 보인 쿼드트리의 비트

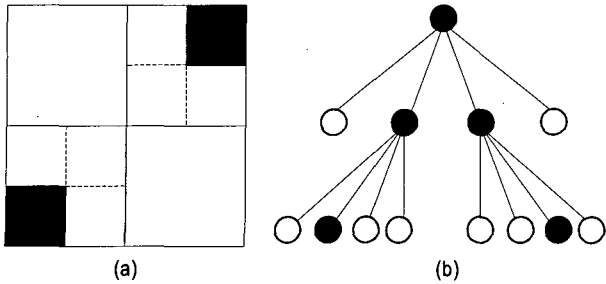


그림 4. SPECK의 예 (a) 중요계수의 위치 (b) 쿼드트리 표현  
 Fig. 4. Example of SPECK (a) Location of Significant Coefficients (b) Quadtree Representation

표현은 1 01 0100 1 0010 0 이 된다. 이때 각 공간들은 SPIHT 에서와 같이 리스트에 저장되어 작은 크기부터 순서대로 부호화를 진행 시킨다. 이 리스트와 쿼드트리 부호화의 효율적인 동작으로 SPECK은 단순한 알고리즘으로 인하여 매우 빠른 동작을 보이면서도, 그림 5에서 보는 바와 같이 동영상 압축에서의 오차 영상과 같이 고주파 영상을 상대적으로 많이 포함하고 있는 영상의 부호화에 있어서는 널리 사용되는 SPIHT 보다 뛰어난 압축 성능을 보인다.

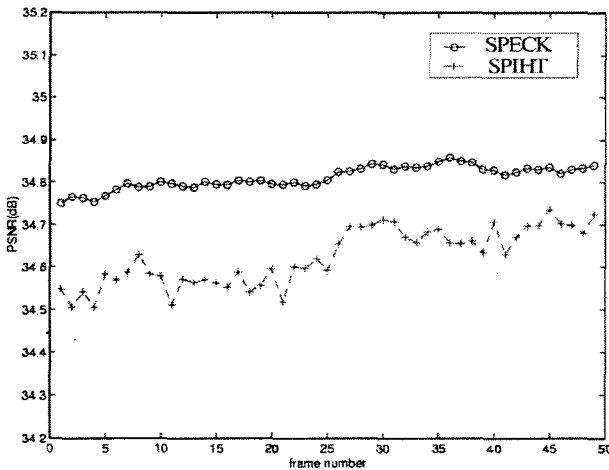


그림 5. SPECK과 SPIHT 비교 (동일 비트율에서 예측 오차 영상 부호화시 PSNR)  
 Fig. 5. Comparison of SPECK and SPIHT (PSNR of reconstructed predicted error frames at the same bit rate)

또한 SPECK의 경우 하나의 대역을 부호화한 이후에 다음 대역에 대한 부호화를 진행하기 때문에, 그림 3에 표시한 EZW 부호화의 경우와는 달리, 주어진 임계값에 대하여

저주파 대역의 비트열이 먼저 발생하고, 차례로 고주파 대역의 정보를 포함하는 비트열이 나타나는 특징을 가지고 있다. 따라서 하나의 비트열로부터 고해상도 표시 장치와 저해상도 표시 장치를 동시에 지원하기에 용이하다. 즉 LL대역에 해당하는 비트열만(그림 6의 흰색으로 표시된 부분)을 추출하여 복원하는 경우 LL 대역의 크기에 해당하는 복원 영상을 재생할 수 있으며, 전체 비트열을 이용하는 경우 고해상도 원본 영상과 동일한 크기의 영상을 복원할 수 있는 것이다. 따라서 SPECK은 다중해상도를 지원하는 동영상 압축 시스템에 채용하기 적합한 부호화 기법으로 생각할 수 있다.

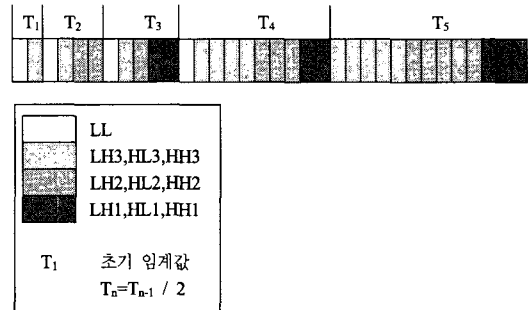


그림 6. SPECK의 경우 비트열의 대역분포  
 Fig. 6. Example of SPECK encoded bit stream distribution along resolution bands

### III. 시간영역 웨이블릿 변환과 움직임 예측

일반적인 동영상 압축은 그림 7과 같이 시간상의 중복을 제거하기 위해 움직임 예측과 보상을 이용하는 예측 부호화 방식과 공간상의 중복을 제거하기 위해 변환기와 양자화기를 이용하는 변환 부호화 방식이 결합된 하이브리드 (Hybrid) 형식을 취한다.

이러한 구조에서는 움직임 예측/보상의 동작이 DPCM의 순환적 (closed-loop) 구조와 본질적으로 동일하며, 전송상의 에러가 발생한 경우 이후 프레임으로 오류가 전파되는 에러 전파(error propagation)가 발생한다. 이것은 근본적으로 예측부호화 방식에서는 피할 수 없는 약점이다. 이에 비해 변환 부호화에서는 에러 전파 현상이 없으므로 전송 에러에 더욱 강한 특징을 가진다. 뿐만 아니라 매우 낮은

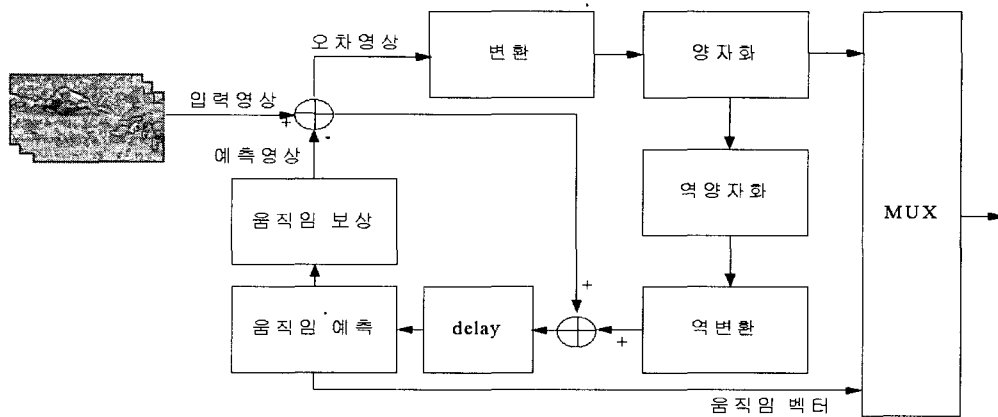


그림 7. Hybrid 동영상 압축의 동작과정  
Fig. 7. Block Diagram of Hybrid Video Encoder

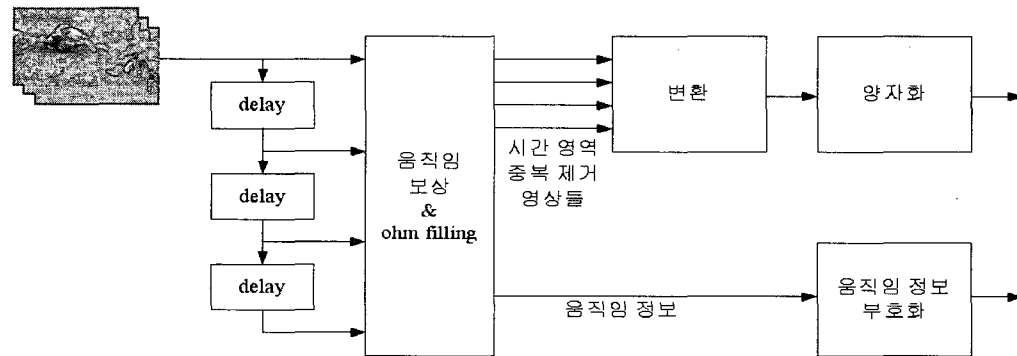


그림 8. MC-3DSBC의 동작과정  
Fig. 8. Block Diagram of MC-3DSBC

전송률로 부호화하는 경우 예측부호화 방식에서는 예측 오차가 전파 될 수 있어 전송률을 낮추는데 있어서 어려움이 있으나 변환 부호화시에서는 이러한 문제 역시 발생하지 않는다. 즉, 웨이블릿 변환을 공간 영역뿐만 아니라 시간영역으로까지 확장한 3D 웨이블릿의 경우 Hybrid 방식의 순환적 구조를 제거해 에러에 대한 안정성을 가지면서 해상도(resolution)와 frame rate 그리고 화질까지 모두 확장성(scalability)을 지닌다. 이로 인해 보다 안정적인 전송과 다양한 기기에 적용이 가능하다는 점에서 향후 높은 발전이 기대되는 네트워크 기반의 VOD 서비스와 같은 분야에서 뛰어난 강점을 가진다. 이러한 3D 웨이블릿 기반의 동영상 부호화 방법으로 잘 알려진 3D-SPIHT<sup>[12]</sup>의 경우 움직임 예측과정 없이 동영상 신호를 일정한 GOP(Group of

Pictures) 단위로 3D 웨이블릿 변환 후 3차원으로 확장된 SPIHT 알고리즘을 이용해 압축한다. 3D-SPIHT은 의료용 단층 영상과 같은 특정 응용분야에서는 매우 우수한 성능을 가지지만<sup>[13]</sup> 일반적인 동영상의 경우에는 움직임 예측 과정이 없기 때문에 그 효율성이 떨어지는 단점이 있다. 이러한 단점을 보완하기 위해서 3D 웨이블릿에 움직임 보상 기법을 적용한 방법에 대한 연구들도 활발하게 전개되고 있다<sup>[14][15]</sup>.

그림 8은 움직임 보상 시간영역 필터링(MCTF: Motion Compensated Temporal Filtering) 기법을 적용한 3D 웨이블릿 압축방법인 MC-3DSBC<sup>[16]</sup>의 블록도이다. MC-3DSBC에서의 핵심적인 부분은 움직임 보상과 Ohm filling 부분으로, 이 부분을 통해 시간 영역의 중복 성분이 제거 된다.

이 부분의 동작은 다음과 같이 이루어진다. 움직임보상과 Ohm filling 부분은 기본적으로 2개의 프레임 A 와 B 를 입력으로 받아 저주파 프레임 L 과 고주파 프레임 H 로 분리 하는 것을 목적으로 하며 이때 A 는 L 로 B 는 H 로 변환 된다. 만약 움직임 보상을 위한 움직임 예측 방법으로 블록 정합법을 사용하고, 그림처럼 A 와 B 가 2개의 움직임 블록으로 이루어져 있을 때 그림처럼 움직임 벡터가 결정 되었다면, A 와 B 의 프레임들은 그림 9와 같은 3가지 영역으로 나누어 볼 수 있다.

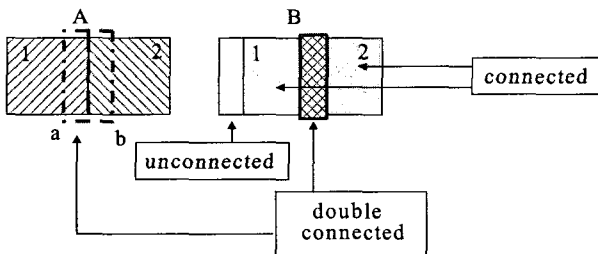


그림 9. 움직임 보상 과 ohm filling 의 예  
Fig. 9. Example of Motion Compensation and Ohm Filling

시간 영역에서의 움직임이 보상된 웨이블릿 변환을 적용하기 위해서는 움직임 예측을 통해 움직임 벡터를 구해서 두 프레임간의 화소들을 서로 연결한 후 이렇게 연결된 화소들 간의 웨이블릿 변환을 적용하면 된다. 하지만 새로이 나타난 물체나 사라지는 물체들 때문에 연결이 되지 못하는 부분이 발생 하게 된다. 만약 A를 기준으로 B에서 움직임 예측을 수행 하였을 때 서로 연결된 부분인 connected 영역, B의 영역에 동시에 연결되는 A 영역들인 double connected 영역, 그리고 B 영역에서 연결이 되지 못한 unconnected 영역이 있다고 하면, 이러한 각 영역 별로 다음과 같은 처리 방법을 사용한다.

**connected 영역 :** A와 B 둘 간의 connected 된 화소들에 웨이블릿 변환을 적용 후 저주파 부분은 A 로 고주파 부분은 B 로 채워 넣는다.

**unconnected 영역 :** B 에 unconnected 영역은 주변 화소들의 움직임 벡터를 참조해 A frame 내에서 reference 화소들을 구한 후 웨이블릿 변환 후 고주파 부분을 채워 넣는다.

**double connected 영역 :** A에서 double connected 된 a 와

b 중 하나를 connected 로 결정해 처리하고 남은 부분은  $\sqrt{2}$  를 곱해 주변에 connected 처리된 주변에 웨이블릿 저주파 계수들과 동일한 에너지 준위를 가지게 한 후 채워 넣는다. 이러한 과정을 거쳐서 A, B 영상을 저주파 계수를 모은 L 영상과 고주파 계수를 모은 H 영상으로 분해할 수 있다. 이렇게 2개의 frame 간의 움직임 보상이 적용된 웨이블릿 변환을 수행 할 수 있게 되면 이것을 N 개의 영상을 가진 GOP 구조로 확장해 웨이블릿 변환에서의 octave 분해 형태를 가지게 할 수 있다.

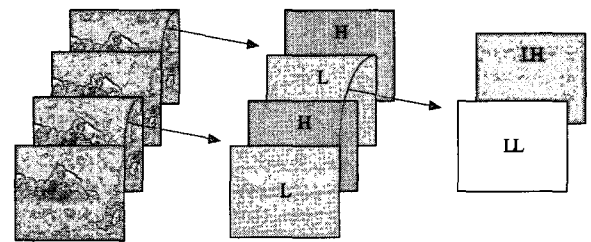


그림 10. 4-GOP 에서 octave 분해  
Fig. 10 Octave Decomposition at 4-GOP

이러한 MCTF 기반 MC-3DSBC 구조는 움직임 보상을 통해 움직임이 많은 동영상에서도 우수한 압축 성능을 가지면서도 다중 재생률도 지원하는 장점이 있다.

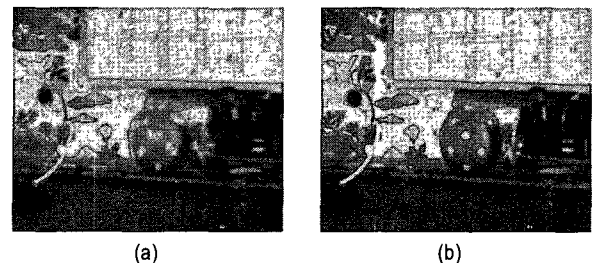


그림 11. 재생률 감소 영상 (a)움직임 보상이 없을 경우 (b)움직임 보상이 적용된 경우  
Fig. 11 Sample frames of frame rate reduced video (a) Without MCTF (b) With MCTF

그림 11에서 보듯이 움직임 보상이 없이 단순히 전후 프레임의 평균값을 사용하여 재생률을 감소시킬 경우 앞/뒤 영상간의 잔상이 크게 남는 단점이 있으나 MC-3DSBC 는 움직임 보상을 이용하므로 이러한 왜곡이 제거된 영상이 얻어진다.

### IV. 다중 블록크기 움직임 예측

제안하는 계층적 구조의 다중 블록 크기 움직임 예측방식 (MBMA; Multi-Block Size block Matching Algorithm)은 최초 크기의 움직임 블록에서 더 작은 크기의 움직임 블록으로 분해해 나가면서 최적의 움직임 블록 구조를 찾는 방식으로 동작하며 일반적인 하이브리드 구조의 웨이블릿 기반 동영상 압축 기법에 적용하였을 때 우수한 성능을 보임을 확인하였다<sup>[8]</sup>. 본 논문에서는 MBMA 기법을 화질 보장형 MCTF 기반 동영상 압축에 적용하였다. 움직임 예측 정보는 움직임 블록의 구조 정보와 각 움직임 블록 별 움직임 벡터의 값으로 구분 된다. 이때 블록의 분해 및 부호화는 SPECK과 유사한 쿼드 트리 분해 및 부호화를 사용한다. 예를 들어 최대 크기의 움직임 블록에서 3 번까지 분해를 허용하며, 그림 12와 같은 결과를 얻었다면, '1'로 해당 블록이 분해되었음을 표시하고 '0' 더 이상 분해되지 않았음을 표시하여 이어서 해당 블록의 움직임 벡터를 첨가하면 된다. 즉 1 1 0(2,0)0(2,1)0(3,1) 1 (4,0)(1,1)(3,2)(0,1) 0(0,-1)0(1,0)0(0,0) 이 출력 비트열이 된다.

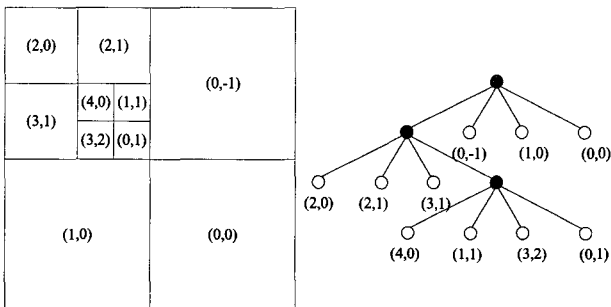


그림 12. 움직임 블록의 정보와 움직임 벡터의 예  
Fig. 12. Example of motion block information and motion vectors

#### 1. Rate Distortion Model

최적의 움직임 예측은 움직임 예측 정보 부호화에 소모되는 전송률  $R(MP)$ 와 예측 오차 영상을 부호화 하는데 소모되는 전송률  $R(RI)$ 의 합으로 표시되는 총 전송률  $R$ 을 최소로 하는 것이다. (여기서 MP는 Motion Prediction, RI

는 Residual Image coding을 의미한다.)

$$\min R = \min [R(MP) + R(RI)] \quad (1)$$

이 때  $R(MP)$ 는 움직임 벡터의 허프만 코드의 크기와 움직임 블록 정보의 합으로 정확하게 구할 수 있지만,  $R(RI)$ 를 구하기 위해서는 우선적으로 사용된 움직임 예측 정보가 주어져야 하며 변환기와 양자화기 과정을 거쳐야만 한다. 웨이블릿 변환은 전체 영상에 적용하는 과정이므로 처음 크기의 움직임 블록에 해당하는 매크로 블록당 소요되는 비트를 정확히 계산하는 것은 따라서 매우 어려우며, 또한 이러한 과정을 실제로 실행하는 것은 엄청난 시간을 소비하는 과정이므로, 적당한 모델을 이용해 예측하여야 한다. 널리 적용되는 바와 같이, 움직임 보상된 예측오차 영상이  $\sigma^2$ 의 분산을 가지는 i.i.d Laplacian 분포를 가지는 신호라고 가정한다면, 다음과 같은 왜곡식을 구할 수 있다.

$$D(R) = \frac{\sigma^2}{\pi} 2^{-2R} \quad (2)$$

이제 원하는 화질에서의 허용 왜곡  $D$ 에서 전송률  $R$ 을 구하면 다음과 같다.

$$\begin{aligned} D(R) &= D \\ R &= R(RI) \\ R &= \min \left[ \frac{1}{2} \log_2 \frac{e\sigma^2}{\pi D}, 0 \right] \end{aligned} \quad (3)$$

이렇게 예측된 전송률을  $R(RI)$ 로 간주함으로써 주어진 움직임 예측 정보로 인해 매크로 블록 당 소요될 전송률  $R$ 을 예측할 수 있다.

#### 2. 움직임 예측 과정

MBMA에서 움직임 블록의 구조를 최적으로 하기 위해서는 발생 가능한 모든 움직임 블록 구조를 테스트해보면 (Exhaustive Search)알 수 있다. 하지만 모든 구조를 다 찾

아보는 것은 너무 많은 연산을 요구 하는 것이므로, 현재의 움직임 블록을 분해하기 전과 후에서 총 소요 전송률이 감소하는 경우에만 블록을 분해해가는 Greedy Algorithm을 적용하여 연산량을 감소시킬 수 있다.

$$\Delta R = [\Delta R(MP) + \Delta R(RD)] < 0 \quad (4)$$

즉, 위의 식이 참이면 움직임 블록의 쿼드 트리 분해를 허용하고 거짓이면 해당 움직임 블록은 더 이상 분해 하지 않는 방식으로 동작 하는 것이다.

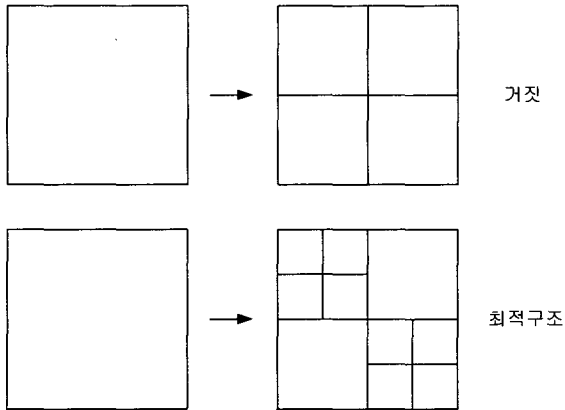


그림 13. 최적 구조를 찾지 못하는 경우  
Fig. 13. Example of failing in finding optimal structure

그림 13에서 보듯이 처음 분해에서는 거짓이었지만 다음 번 분해가 되어서 최적의 움직임 블록 구조를 가지는 경우, 이러한 처리 과정으로는 최적 구조를 찾을 수 없다는 단점이 있지만 이러한 경우는 자주 발생 하지 않았다.

### 3. 고속처리

움직임 벡터를 구하기 위한 검색과정은 SAD(Sum of Absolute Difference)를 이용하였다.

$$SAD = \sum_x \sum_y [X(x, y, t) - X(x + \Delta x, y + \Delta y, t - 1)] \quad (5)$$

일반적으로 사용되는 16x16 크기의 고정 블록의 움직임

예측에서 반 화소 정밀도의 검색에 총 49번의 블록 정합 연산을 수행 하는 경우에 블록별 12,544 번의 빠기 연산을 수행해야 한다. 제안하는 MBMA 의 경우 움직임 블록의 분해가 없을 때에만 같은 복잡도를 지니고, 나머지 분해 과정과 분해가 합당한지 판별하기 위한 과정들이 오버헤드가 된다. 이는 동영상 부호화에서 가장 비싼 비용의 연산인 움직임 예측 과정이 더욱 복잡해졌음을 의미한다. 이 문제를 해결하기 위해 다음과 같이 가정을 적용하였다. 만약 넓은 움직임 블록이 분해가 일어나지 않았다면 넓은 공간이므로 다운 샘플을 적용해도, 즉 해당 블록내의 일부의 화소만을 이용하여 SAD를 계산하여도, 움직임 벡터를 찾아 낼 수 있을 것이다. 이와 같이 블록을 가로세로 2:1로 다운샘플 하여 SAD 를 구할 경우 연산량은 1/4로 감소한다. 뿐만 아니라 배경과 같이 움직임이 없어 움직임 예측이 잘 수행된 경우 R(RI) 는 0 보다 작은 값으로 예측 되며, 단 한 번의 테스트로 움직임 블록 구조와 벡터의 값이 구해지므로 고정 블록의 경우 보다 더 적은 연산량으로 움직임 예측 정보가 구해진다. 이렇게 첫 번째와 두 번째 분해의 비교적 큰 크기의 움직임 블록에서는 가로세로 각각 2:1 다운 샘플한 SAD 를 이용하고 이보다 더 작은 크기로 분해되어졌을 때는 블록내의 모든 화소를 이용한다. 즉 예측이 쉽게 될 경우는 거칠게 그렇지 않을 경우 정밀하게 블록 정합을 수행 하여 연산 비용을 균형 있게 분배하게 된다.

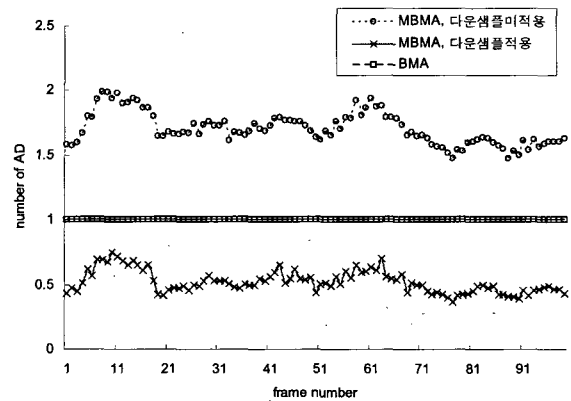


그림 14. 고속처리를 적용한 MBMA 와 BMA 연산량(AD의 회수) 비교 (Salesman 영상)  
Fig. 14. Computational complexities (number of absolute differences) of high speed MBMA and regular BMA (Salesman)



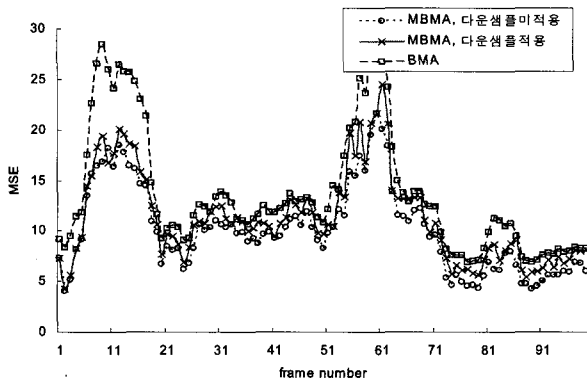


그림 15. 고속처리를 적용한 MBMA 와 BMA 의 MSE 비교 (Salesman 영상)  
 Fig. 15. MSE values of high speed MBMA and regular BMA (Salesman)

이와 같이 제안한 MBMA 에서는 움직임 블록의 구조에 따라 개별 움직임 블록의 총 연산량이 변화하게 되므로 한 프레임 내에서 총 연산량 역시 가변적이다. 그림 14, 15에 보인바와 같이 모의실험을 통해 테스트 동영상의 성질에 따라 조금씩 다르기는 하지만 고정 블록을 사용할 때에 비해서 40~50% 정도 수준의 연산만을 필요로 하였으며 예측 성능은 개선됨을 확인할 수 있었다.

4. SPECK와 MBMA의 확장

본 논문에서는 MBMA구조를 Ohm Filing이 적용된 MCTF 동영상 압축에 적용하고 [8]의 경우와 마찬가지로 최저 화질 보장형 가변 비트율 구조의 동영상 압축 기법을 제안한다. MBMA 와 MCTF 과정을 거치면 최종적으로 3 차원 웨이블릿 계수 들 만이 남게 되며 이 웨이블릿 계수를 부호화하기 위해 3차원으로 확장된 SPECK를 이용한다. 이때 확장된 SPECK은 GOP 내에서의 전체 MSE 을 최소로 하는 것을 목적으로 한다. SPECK 부호화기는 별도의 전송률제어(Rate Control) 없이도 주어진 비트 내에서 최적의(혹은 최적에 가까운) 부호화가 가능한 부호화기이다. 때문에 부호화의 정지를 일정 비트를 모두 소모하는 것에서 주어진 화질에 도달할 때로 바꾸는 것으로 주어진 화질에서 최적인 압축이 가능하다. 만약 GOP의 원하는 평균 화질이 PSNR Q dB라면, 그에 따르는 MSE를 아래 식과 같이 구할 수 있다.

$$PSNR Q = 10 \log_{10} \left( \frac{255^2}{MSE} \right)$$

$$MSE = \frac{255^2}{10^{Q/10}}$$

$$SE = MSE \times \text{Number of Wavelet Coeff.}$$

부호화 과정에서 아직 부호화 되지 못한 계수의 에너지의 총합이 SE와 같아지면 부호화를 멈춘다. 이것으로 GOP 내에서 주어진 평균 화질에 도달하는 최적의 비트열을 구할 수 있다.

또한 2개의 프레임을 MBMA 와 Ohm filling 시 H 프레임만이 바로 출력되는 프레임이며 또한 [8]의 Hybrid 형태의 부호화기 에서의 P 나 B 프레임에 해당 하는 것이므로 [8]에서 적용한 R(RI) 예측 과정을 그대로 H 프레임에 적용해 사용할 수 있다.

V. 모의실험 및 결과

제안 방법은 주어진 화질에 최적으로 비트를 할당하여 압축을 수행한다. 이때 화질은 영상에서의 화질을 표현하는 객관적인 기준인 PSNR 을 이용하였다. 또한 제안 방법의 효율성을 비교하기 위해 H.263 부호화기와 동일한 화질로 압축을 수행하였을 때 필요로 하는 비트를 비교 하였다. 이때 제안 방법은 4-GOP 구조로 동작하게 하였으며, MBMA 는 64x64 크기의 초기 움직임 블록에서부터 2 번 까지 분해를 허용 하게 하였으며, 움직임 벡터는 H.263 과 동일한 허프만 테이블을 사용하여 부호화 하였다. 비교 대상인 H.263 은 제안 방법과 같이 4-GOP 구조로 동작 한다. 실험 영상은 gray-scale 의 Susie 영상으로 MBMA 초기 움직임 블록 크기가 64x64 로 설정되어 있어 64의 배수 크기의 넓이와 높이를 가진 것으로 사용하였다.

표 1. Susie 영상에서 평균 화질과 GOP 별 전송률  
 Table 1. Average Quality and bit rates per GOP for Susie

	PSNR(dB)	출력크기(bit)	전송률(bpp)
H.263	36.07	25731	0.082
제안방법	35.89	10500	0.033

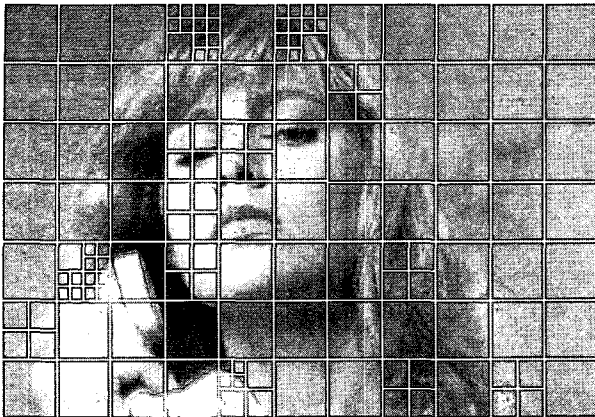


그림 16. MBMA 에 의해 분할된 움직임 블록  
Fig. 16. Motion block determined by MBMA



그림 17. Susie 영상 14번째 프레임 복원 화면 (PSNR=36.14 dB)  
Fig. 17. Reconstructed 14-th frame of Susie (PSNR = 36.14 dB)

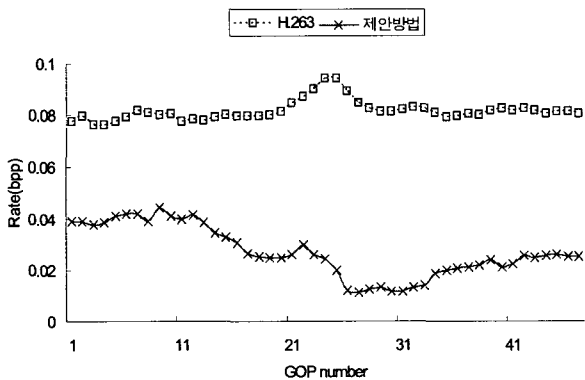


그림 18. Susie 영상에서의 비트소모량 비교  
Fig. 18. Bit rates for Susie sequence

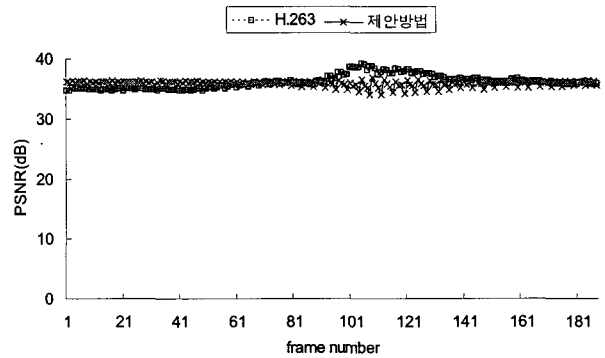


그림 19. Susie 영상에서의 화질 비교  
Fig. 19. Picture quality of Susie sequence

## VI. 결론 및 향후 과제

MBMA는 요구되는 연산 복잡도가 낮음에도 불구하고 움직임 예측 정보 부호화에 소요되는 비트와 예측오차 프레임의 웨이블릿 변환 계수 부호화에 소요되는 비트 간에 합리적인 비트 할당을 해주며, SPECK은 웨이블릿 계수의 부호화시 GOP 단위로 주어진 평균 화질에 최적 비트 할당을 가능하게 해준다. 제안한 MCTF기반 동영상 압축 방식은 H.263 에 비해서 약 60% 정도 우수한 압축 성능을 보였다. 보다 최신의 H.264와 비교하면 압축 성능의 우월성은 부족할 것으로 예상되나, 압축 과정의 복잡도를 고려하면, 제안하는 기법도 우수한 압축 기법이라고 할 수 있다. 또한 시간 방향의 중복 성분 제거에 순환적 구조(Closed loop)의 예측 부호화기가 아닌 MCTF 기법을 사용하였기 때문에 다중 재생률 지원과 전송 오차에 강인한 장점도 가지고 있다. 제안 방법의 동영상 압축방식이 가진 이러한 성질은 향후 네트워크를 통한 VOD 서비스 등에 사용되기 적합한 방식이라 하겠다.

## 참고 문헌

- [1] ISO/IEC 11172-2 International Standard of MPEG-1, Part2: Video, 1993.
- [2] S.-Y. Huang and J.-S. Wang, "A low-cost desktop video conferencing codec: an adaptive motion-JPEG design," IEEE Trans. on

Consumer Electronics, Vol. 40, No. 4, pp. 944-950, Nov. 1994.

[3] ITU-T Recommendation H.263: Video Coding for Low Bit Rate Communication Mar. 1996.

[4] V. Bhaskaran and K. Konstantinides, "Image and Video Compression Standards - Algorithms and Architectures", Kluwer Academic Publisher, 1995.

[5] K. Rao and P. Yip, "Discrete Cosine Transform-Algorithms, Advantages and Applications", Academic Press, London, UK, 1990.

[6] Y.T. Chien and K.S. Fu, "On the generalised karhunen-loeve expansion", IEEE Tran. on Information Theory, IT-13, pp. 518-520, 1967.

[7] K. Shen and E.J. Delp, "Wavelet based rate scalable video compression", IEEE Trans. on Circuits Syst. Video Tech., Vol. 9, No. 1, pp. 109-122, Feb. 1999.

[8] Sang Ju Park, Hyoung-Jin Kim and Min Chul Park, "A Rate Control Scheme Using Multi Block Size BMA for DWT-Based Video Compression with Constant Quality", IEICE Tr. on Fundamentals, Vol.E87-A, No. 6, pp. 1426-1432, June. 2004.

[9] A. Islam and W. A. Pearlman, "An Embedded and Efficient Low-Complexity Hierarchical Image Coder," Proceedings of SPIE Visual Communications and Image Processing, Vol. 3653, pp. 284-305, January 1999.

[10] Jerome M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", IEEE Trans. on Signal Processing, Vol. 41, No. 12, pp. 3445-3462, Dec. 1993.

[11] A. Said and W. A. Pearlman, "A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees", Trans. on Circuits Syst. Video Tech., Vol. 6, No. 3, pp. 243-250, Jun. 1996.

[12] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3D set partitioning in hierarchical trees (3D SPIHT)", IEEE Trans. on Circuits and Syst. for Video Tech., Vol. 10, No. 8, pp. 1374-1387, Dec. 2000.

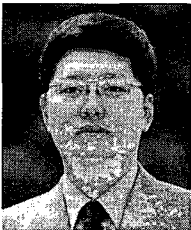
[13] W. A. Pearlman, "Trends of Tree-Based, Set Partitioning Compression Techniques in Still and Moving Image Systems," Proc. PCS-2001, pp. 1-8, Apr. 2001.

[14] F. Bosveld, R. L. Lagendijk, and J. Biemond, "Hierarchical video coding using a spatio-temporal subband decomposition", Proc. ICASSP-92, Vol. 3, pp. III/221-III/224, Mar. 1992.

[15] M. P. Quefuz, "A 3-dimensional subband coding scheme with motion-adaptive subband selection", Proc. EUSIPCO-92, pp. 1263-1266, Sep. 1992.

[16] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," IEEE Trans. on Image Proc., Vol. 3, No. 5, pp. 559-571, Sep. 1994.

저 자 소 개



박 상 주

- 1984년 : 서울대학교 전자공학과 (학사)
- 1986년 : 서울대학교 전자공학과 (석사)
- 1992년 : Stanford University, Electrical Eng., (박사)
- 1992년~1993년 : Philips Research Palo Alto, MTS
- 1993년~1995년 : Philips Semiconductors, MTS
- 1996년~현재 : 홍익대학교, 전자전기공학부 교수
- 주관심분야 : 영상 처리, 디지털 워터마크, 정보압축