

# 웨이블릿 기법을 적용한 저속 비트율 비디오 코딩에 관한 연구

김 혜경<sup>o</sup>, 이 옥경, 지 정규, 오 해석  
숭실대학교 컴퓨터학과 멀티미디어연구실

## A Study of Very Low Bit-Rate Video Coding Using Wavelet-Based Techniques

Hye\_Kyung Kim, Ok\_Kyung Lee, Jeong\_gyu Jee, Hae\_Seok Oh  
Department of Computer science, Soongsil University

### 요 약

본 논문에서는, 이산 웨이블릿 변환(DWT: discrete wavelet transform)에 기초한 저속 비트율 비디오 코딩 방법의 알고리즘으로써, PZTACC 방법을 제안한다. 접근 방법은 양자화된 웨이블릿 계수들이 웨이블릿 서브밴드 구조내에서 중복성을 활용하는 메커니즘에 의해서 전처리 된다면 코딩 절차가 더욱 효과적으로 작용하여 나타난다. 그러므로 본 논문에서는 partitioning, Zero-Tree aggregation, 조건적 코딩의 개념들을 기초로 하는 프리코딩 기법의 새로운 구조를 소개한다.

복원된 비디오의 주관적인 성능은 일반적으로 MPEG4의 VM 구현으로부터 획득된 성능(1.5 dB PSNR 이상)보다 우수하다. 게다가, 인트라프레임모드에 제한되어 있을 때, 제안된 코딩 알고리즘은 정지영상 압축에 대하여 최고로 보고되는 결과를 산출한다.

### 1. 서론

매우 낮은 비트율의 비디오 압축은 저대역의 채널과 함께 작동하는 비디오 통신시스템을 빚어 말하고있으며, 빠르게 확산되는 연구 분야가 되었다. 게다가 잘 확립된 기법들은 블록기반 움직임 보상(BMC:block-based Motion compensation)과 DCT를 기초로 하고 있으며, 잘 알고 밝혀진 H.263 표준안[4]내에서 파산되었으며, 새로운 아이디어와 기법들이 드러났다.

이러한 기법들 중에서 가장 유망한 한가지는 웨이블릿 변환 스킵이며, 이런 변환 코딩기법은 부대역(sub-band) 코딩 기법들로 결합된다. 웨이블릿은 현상 공간내(즉, 공간 주파수 도메인) 잘 배치되고, 영상의 자연적인 특성들에 부합되어, 그들의 다양한 규모의 특징들이 드러난다. 웨이블릿 기반 코딩 스킵의 변형 모듈은 일반적으로 양자화에 의해 수행되고, 시각적으로 관련없는 정보를 무시하고, 그리고 마지막 단계에서는, 엔트로피 코더가 양자화에 의해 발생된 데이터 스트림내에서 통계적인 중복성을 제거하기 위해 사용된다. 이런 일반적인 웨이블릿 기반 코딩 스킵은 정지영상이나 비디오에 적용될 수 있고, 최근에는 정지영상이나 비디오에 시간적인 중복성을 이용하는 메커니즘을 추가하는 것이 제공되고 있다.

이러한 작업에 있어서, 본 논문에서는 코딩부분에 대한 최적화 활동에 초점을 맞추고, 즉 엔트로피 코딩보다 우선하는 양자화된 웨이블릿 계수들을 전처리하는 소위 프리코딩 기법들의 프레임워크를 제안한다. 제안한 프리코딩 프레임워크는 파티션닝(Partitioning), 집합체(aggregation), 그리고 조건적인 코딩(conditional coding) 개념들을 기초로 하고 있다. 본 논문에서의 아이디어는 웨이블릿 변환의 지역적 특성에 대한 지식을 따르고 있으며, 웨이블릿 표현의 특성들로부

터 생긴 중복성을 유효하게 이용하는 메커니즘을 사용한다.

본 논문의 구성은 다음과 같다. 2절에서, PZTACC 비디오 코더에 대하여 제안된 주요 개념을 설명한다. 3절에서는 PZTACC 알고리즘을 기술한다. 4절에서는 산술적 코딩을 설명하고, 마지막 결론 및 향후 연구방향을 5절에서 나타낸다.

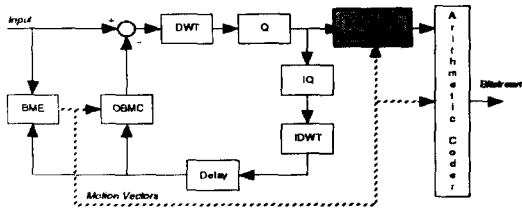
### 2. PZTACC-부호기(Codec)의 개요

부호화 알고리즘의 단순화된 블록 다이어그램은 [그림1]에 주어졌다. 그것은 본질적으로 3개의 부분으로 구성한다. 첫 번째 부분은 시간적인 예상 피드백 루프로써, 예상(prediction)은 블록 매칭 추정(BME:black-matching estimation)과 OBMC를 사용하여 수행된다[1]. 이러한 순환 프로세싱의 초기 상태는, 소위 인트라프레임 모드(intraframe mode)라 불리고, 예언적인 부분은 직접적으로 이산 웨이블릿 변환(DWT)으로 들어가 우회한다.

양자화(Quantization)한 후에, 프리코더(precoder)는 최종적인 산술 코딩 단계에서 웨이블릿 변환 고유의 중복성을 효과적으로 이용하는 것을 허용하기 위해서 양자화된 웨이블릿 표현의 전처리를 수행한다. 역양자화(IQ:dequantization)와 역DWT(inverse DWT:IDWT)는 인트라프레임(interframe) 즉, 예언적인 모드(predictive mode)내에서 다음 입력 프레임의 처리를 위해 재구성된 프레임으로 예언적 루프(loop)를 만족시킨다.

### 3. PZTACC Precoding 알고리즘

코딩의 효율성에서 더 나은 개선점을 달성하기 위해서, 더 많이 복잡하고 정교한 모델이 설계되어야 한다. 다소간의 높은 통계적 의



[그림 1] 부호화 알고리즘의 블록 다이어그램

존성의 정도를 주어진 데이터에서 추정하여, 순서 N의 모델로 간단한 접근, 예를 들면, 주어진 알파벳 크기인 M 으로 랜덤 벡터  $s = (s_0, \dots, s_N)$  의  $M^{N+1}$ 개의 다른 값들에 대한 확률분포 (PD)를 평가하도록 요구하였으며, 그것은 충분히 큰 N과 M에 대해서, 일반적으로 실행 불가능하다. 그러므로 다소 단순하게 만들어야 하고, 반면에 통계적인 의존성의 정도를 최대한 유지한다.

이러한 방침에서 첫 번째 단계는 파티션(partitioning)에 관한 도구를 가지고 소스를 나누는 것과 소스에 대한 알파벳 크기 M을 줄이는 것이다. [그림2]에서 PZTACC 프리코딩 구조에 대한 최초의 파티션 과정은 세 개의 서브소스로 끝나는 것을 설명한다. 즉, 의미도는 의미 있는 계수들의 위치를 나타내고, 크기도(magnitude map)는 중요한 계수들의 절대적인 값을 저장하고, 그리고 부호도(sign map)는 웨이블릿 계수들에 관한 단계적 정보를 저장한다.

세가지 서브소스들 모두 웨이블릿 분해로부터 부대역(subband)의 구조를 상속하고, 그결과 각 서브소스로부터 2가지 대역 구조에 의하여 다른 파티션을 얻는다. 프리코딩 방법의 두 번째 단계는 소위 제로-트리 데이터 구조[2]를 사용하여 다른 대역을 교차하여 대수롭지 않은 계수들의 집합체로 구성한다.

프리코더(precoder)[1]의 주요한 부분은 "내용(context)" 즉, 산술 부호기에서 실제적인 코딩 과정에 대한 적당한 모델을 도구로 각 소스의 구성요소를 공급한다.

1) Partitioning : 분할의 이론상 기초는 두가지 원리에 의해 주어지고, 웨이블릿 기반의 이미지 코딩에서 유용한 것으로 발견되었다.

첫 번째 이론은 소스에 대한 엔트로피 비율은 소스를 빈공간이 아닌 서브소스들로 해체하여 나누는 것에 의하여 축소될 수 있다는 것을 말한다. 이러한 소스 분할의 원리는 한가지 명백한 애플리케이션을 갖는다. 즉, 양자화되고 웨이블릿 변환된 이미지 c를 부대역 구조에 따라 서브소스  $C_{i,k}$  로 다른 확률분포와 함께 나눈다. 비록 이러한 분할의 첫 번째 단계가 이론상의 엔트로피 비율을 이미 감소시켰을지라도, 적용할 수 있는 범위에 대한 반복되는 단계의 두가지 분할을 추가

하고, 다음과 같은 이론에 의해 특성화된다.

이론1(Adaptive Range Partition) : 소스 S에 대한 동적인 범위가  $A = \bigcup_i A_i$  에 의해 주어졌고, 거기에서  $A_i$  는 해체하고, 빈공간이 아닌 A의 부분집합(subsets), 그리고 서브 소스인  $S_i = \{s \in S | s = \alpha, \alpha \in A_i\}$ 를 정의하고, 표시기 집합  $\chi = \{x_k | x_k = i, \text{ if } s_k \in A_i\}$  즉, 그 다음에 전체 엔트로피 비율은 다음에 의해 주어진다.

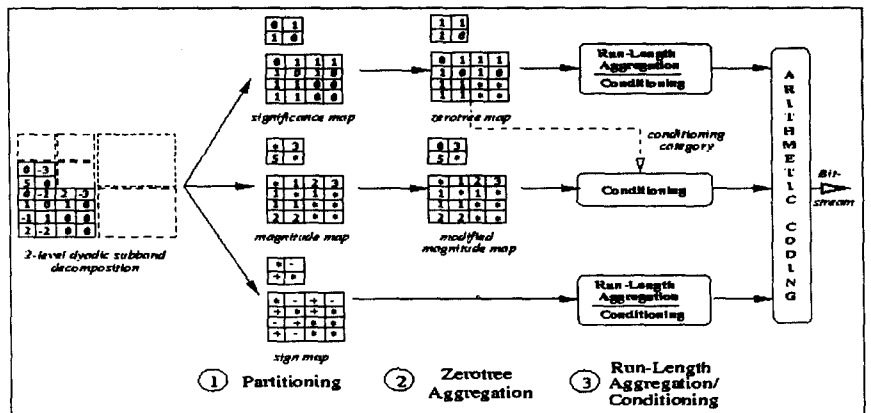
$$R(S) = \sum_i R(S_i) + R(\chi) \quad (2)$$

적용적인 범위의 분할은 엔트로피 비율이 증가하지 않고, 그것은 정보가 영커있는 것을 풀어내도록 허락한다. 제로가 아닌 것으로 양자화된 의미있는 값과 제로로 양자화된 값을 가지는 의미없는 값으로 범위를 나누는 것은, 의미있는 계수  $c_{i,k}^{mag}$  에 대한 서브소스를 초래할 것이고, 그리고 소스는 적용적인 분할에 대한 측면정보를 구성하고, 소위 의미도(significance map)  $x_{i,k}$  라고 부른다. 더 나아가서 범위분할은 의미있는 계수들의 부호에 따라서 마침내 (2)를 사용하여 산출한다.

$$R(c_{i,k}) = R(x_{i,k}) + R(c_{i,k}^{mag}) + R(s_{i,k}) \quad (3)$$

그리고 거기에서 크기도(magnitude map)  $c_{i,k}^{mag}$  는 의미있는 계수들의 크기를 구성하는 서브소스이고 부호도(sign map)  $s_{i,k}$  는 관련된 부호정보를 유지한다.

등식(3)은  $c_{i,k}$  가 두 개의 이진수로 평가된 표시도  $x_{i,k}$  와  $s_{i,k}$  로 교체되고, 크기도  $c_{i,k}^{mag}$  는 도달할 수 있는 코딩 비율에 대하여 더 낮은 한계로 변경할 수 없고, 그러나, 더욱 중요한 것은, 크기도는 이러한 궁극적인 경계에 접근할 수 있는 방법으로 통찰력을 부여한다. 파티션하는 과정 자체는 추상화에 대한 더 높은 수준의 결과이므로, 그것은 다른 서브소스들의 상호의존성에 대한 보다 나은 이용을 허락하고, 다른 타입에 대한 것이든지 또는 다른 주파수 내용에 대한 것이든지, 다음과 같은 두가지 단계가 있다.



[그림 2] PZTACC 알고리즘

2) Zero-Tree Aggregation : 매우 낮은 비트율의 비디오 코딩에 대하여, 다음의 두 가지 요구사항은 필수적이다.

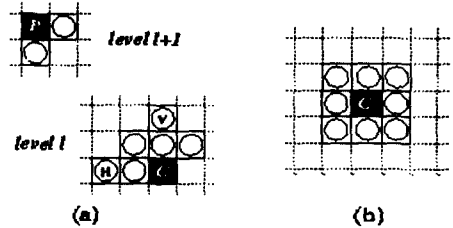
- ① 인트라 프레임(I frames)의 양자화 후에, 유일한 계수들의 작은 파편이 제로가 아닌 값으로 남아 있을 것이다.
- ② 시간적인 예언 구조에 대한 출력은 낮은 에너지에 대한 예상 에러 신호(prediction error signal)일 것이며, 제로가 아닌 계수들이 거의 없는 것으로 양자화 된 웨이블릿 표현이 끝난다.

제로-트리 데이터 구조를 사용한 접근방법은 주로 두 가지 면에서 이전의 연구법과 다르다. 첫째, 제로-트리 루트 심볼을 다루는 방법은 다른 제안들과 같지 않다. 주어진 의미도  $x_{l,k}$  ( $l > 1$ )는, 본 논문에서는 각 무의미한 계수에 대하여 ZT(Zero-Tree) 루트이다. 두 번째 이면서 다른 제로-트리-기반의 방법과 구별되는 주요한 차이는 제로-트리 도구의 사용과 관련된다. 조건적인 코딩과 관련하여 제로-트리 기반의 효율성에 대한 주의 깊은 실험적인 조사는 대역  $W_{l,k}$  에서 최대 분해 수준  $l < l_{max}$  아래로 제로-트리 루트 심볼을 사용하는 것에 대한 더 나은 이득이 없음을 보여준다. 저주파수에 대한 부대역  $l = l_{max}$  로 제한하고 있는 제로-트리 코딩은, 인터프레임 모드에서 대역 내의 루트 심볼을 인덱스  $k=1...3$  로 연결함에 의해 가장 낮은 주파수 대역 ( $k=0$ ) 내에서 같은 공간적 위치와 관련된 제로 계수에 대한 코딩 효율성을 더욱 더 향상시킨다.

3) Conditional Coding : 이전 값으로 된 표시맵(indicator map)에 대한 인코딩을 위하여, 2단식의 이미지에 대한 (무손실) 압축을 위해 개발된 모든 종류의 코딩 방법을 사용할 수 있다. 실행-길이 코딩(Run-length coding)은, 커다란 범위에 대해 2-D 상관 관계를 획득하는 것이 불가능한 이래 한정된 압축 가능성을 가지고 있는 하나의 가능성이다. 양자택일로, 조건적인 확률을 사용한 모델을 기초로 하는 접근 방법은 조절하는 "내용(context)"은 소위 템플릿이라 불리는 것의 도움으로 창조된다. 템플릿은 일반적으로 인코딩하기 위해서 현재의 구성요소와 인접한 요소를 맞추어준다.

[그림 3(a)]는 제로트리 맵  $x_{l,k}^{ZT}$  의 더 낮은 단계 ( $1 < l_{max}$ ) 의 코딩을 위한 목적으로 설계한 템플릿을 보여준다. 이것은 JBIG 표준에서 사용된 미분계층(차분계층)의 템플릿과 유사하고, 두 계층의 구성요소가 덮고 있는 즉, 현재의 요소인 (C)를 둘러싼 5개의 구성요소들, 그리고 C에 대한 "parent" P 의 두 이웃들, 그것은 C의 원인이 되지 않는 근접함에 대한 "prediction"을 허락한다. 게다가, 대역의 방향 k에 대하여 유력한 상관관계에 대한 방향에 따라서 구성요소들 (V, H)중의 한가지를 선택함에 의해 템플릿을 적용한다 [그림 3(a) 참조].

[그림 3(b)]는 현재의 의미있는 계수  $c = c_{l,k}^{m,m}[n, m]$  에 대한  $3 \times 3$  윈도우를 보여주고 있으며, 그것은 c에 대하여 조절하는 기술자  $k = k_{l,k}[n, m]$  를 정의하기 위하여 제로-트리 맵의 부분과 대응하여 "연결된" 것이다. 즉, 적용적인 산술 코더에서 종합적인 학습 비용을 줄이기 위하여, 5개 상태로 특성화된 범주를 조절하기 위한 수를 제한하고 그곳에서 최고의 성능을 산출한다.



[그림 3] (a) 7개의 요소로 된 템플릿 (b)  $3 \times 3$  윈도우

#### 4. Arithmetic Coding

프리코더(precoder)에 의해 다루어진 움직임 벡터와 모든 심볼들은 적용적인 산술적 코더(AAC)를 사용하여 인코딩된다. 실제적으로, AAC를 이전 알파벳으로 제한하는 것에 의해 주어진 실행에 대한 편차를 사용한다. 계수들이나 움직임 벡터의 크기같은 다중 알파벳 심볼들은 그들의 기대된 확률분포에 비례하는 길이를 가진 이전 심볼로 처음 연결되고, 그 때문에 실제적인 통계에 대한 모델들의 더 빠른 적용을 허락한다.

인트라프레임과 인터프레임 모드에 대하여, 독립적인 모델을 사용한다. 연속적인 움직임 벡터 분야 뿐 아니라 연속적인 P 프레임도 이전의 P프레임과 움직임 벡터 분야 각각에 대해 갱신된 관련모델을 사용하여 인코딩된다.

#### 5. 결론 및 향후 연구방향

본 논문에서는 비디오와 정지영상 코딩 알고리즘을 분할(partitioning), 제로-트리 집합체(aggregation), 그리고 조건적인 코딩(conditional coding) 개념을 포함하는 새로운 프리코딩 전략으로 PZTACC 알고리즘을 소개했다. 부호기는 유일하게 프리코딩(precoding) 부분에서 최적화되고, 즉, 표준 DWT를 사용하고, 간단한 불변의 양자화를 사용하여, 전통적인 산술적 코더를 약간 수정된 표준 움직임 압축 기법과 마찬가지로 잘 사용한다. 인트라프레임/정지 영상 코딩과 비디오 코딩에 대한 두가지 의미의 비트율과 비교한 특성에서 이러한 웨이블릿 기반 코더가 고도로 효율적이라는 것을 보여주었다. 향후의 연구 방향은 적용적인 웨이블릿(packet) 변환과 더 나은 움직임 모델의 결합에 대하여 연구가 이루어질 것이다.

#### References

- [1] D.Marpe and H.L.Cycon, "Efficient pre-coding techniques for wavelet-based image compression," in Proc. Picture Coding Symp. 1997, pp.45-50.
- [2] S.A. Martucci, I. Sodagar, T. Chiang, Y.Zhang, "A zerotree wavelet video coder," IEEE Trans. Circuit Syst. Video Technol., vol. 7, pp. 109-118, Feb. 1997.
- [3] T.Sikora, "The MPEG-4 video standard verification model," IEEE Trans. Circuits Syst. Video Technol., vol 7, pp.19-31, Feb. 1997.
- [4] ITU-T Draft Recommendation "Video coding for low bitrate communication," Dec. 1995.