

# 분산 비디오 압축 기술

심현재·전병우 (성균관대학교)

## I. 서론

최근에 휴대 단말이 점차 고성능화 되어감에 따라 정지영상에 대한 부/복호화 기능이 이제 거의 기본 기능으로 탑재되더니 이젠 UCC (User Created Contents) 등이 등장하면서 이제는 휴대 단말을 이용한 동영상 부호화에 대한 요구가 커지고 있다. 그러나 H.264/AVC 등과 같은 기존 비디오 부호화 압축 기술은 1: N의 방송 응용에 적합한 “고복잡도 부호화와 저복잡도 복호화”라는 구조를 가지기 때문에, 방송응용을 염두에 두고 개발된 움직임 보상 변환 부호화(MC-DCT)와 같은 기존의 부호화 기술은, 부호화 알고리즘의 복잡도 및 전력소모 면에서 휴대 단말에는 적합하지 않다.

분산 비디오 압축 기술(Distributed Video Coding)은 바로 이러한 문제를 해결할 수 있는 새로운 기술적 가능성을 제시하기 때문에 최근 들어 많은 관심을 받기 시작하고 있다. 이 기술은 그 구조상, 부호화기의 경량화를 가능하게 하기 때문에 휴대용 단말이나 센서망 등에 활용하기에 적당하다. 이런 이유로 최근의 분산 환경 상의 응용에 적합한 기술로 부상하고 있다. 그러나 아

직은 그 기술의 완성도가 충분히 높지 않아 추가로 연구 개발해야 할 문제들이 산적하여 있으나, 향후 이러한 기술적 문제가 해결되어 움직임 보상 변환부호화 구조 하에서의 지금의 최고 정점 기술인 H.264/AVC 정도의 부호화 효율을 갖게 되거나 실용적으로 사용될 만큼 충분히 성능이 향상된다면, 향후 다양한 응용수요나 이 기술이 주변에 미치는 파급 효과가 매우 클 것으로 예상된다.

본고에서는 현재까지 개발된 분산 비디오 압축 기술을 소개하고, 현재 이 기술의 성능과 몇 가지 문제점 등을 살펴본다.

## II. 분산 비디오 압축 기술

### 1. 이론적 배경

분산 비디오 압축 기술의 이론적 배경은 1970 년대에 발표된 Slepian-Wolf 이론이다. 식 (1) 과 <그림 1>에서 보는 바와 같이, Slepian-Wolf 이론은 두 개의 통계적으로 상관된 신호 X 와 Y를 독립적으로 부호화하여도, 두 상관된 신호를 결합복호화(joint decoding) 한다면 두 신

호 X와 Y를 결합부호화(joint encoding)하는 경우와 동일한 부호화효율을 얻을 수 있음을 수학적으로 증명한 것이다<sup>[1]</sup>.

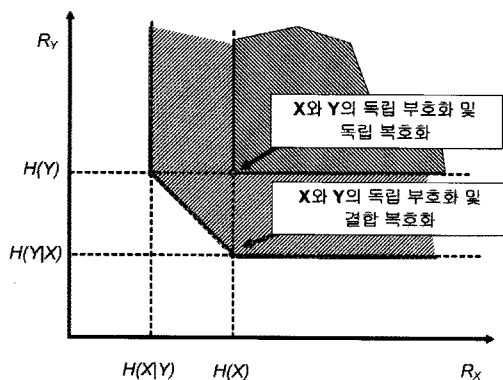
$$R_x + R_y \geq H(X, Y), R_x \geq H(X|Y), R_y \geq H(Y|X) \quad (1)$$

Slepian-Wolf 이론이 무손실 압축의 경우를 가정한데 반해, Wyner-Ziv 이론은 이를 손실 압축의 경우로 확장할 경우에도 유사한 결과가 얻어짐을 보였다. 즉, X와 Y가 결합 가우시안(jointly Gaussian)이면 손실 압축의 경우에서도 Slepian-Wolf 이론을 만족할 수 있음을 의미한다<sup>[2]</sup>. 여기에서 왜곡치는 MSE(Mean Squared Error)로 측정됨을 가정한다.

$$R^*(d) \geq R_{X|Y}(d) \quad (2)$$

여기서  $R^*(d)$ 는 보조정보 Y가 복호화기에만 알려져 있을 경우의 율-왜곡 함수이며,  $R_{X|Y}(d)$ 는 보조정보 Y가 부호화기와 복호화기에 동시에 알려져 있을 경우, 즉 결합부호화(joint encoding)할 경우의 율-왜곡 함수이다.

이러한 정보이론 관점에서의 수학적 결과가

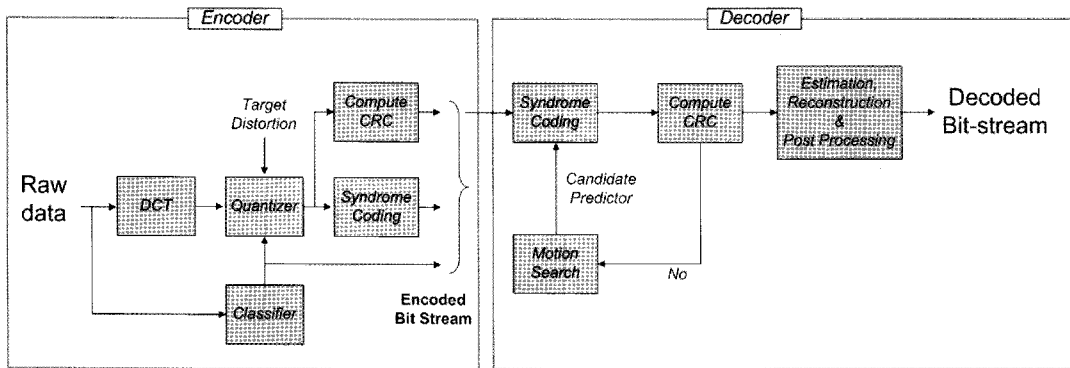


〈그림 1〉 압축가능(achievable) 비트율 영역

오래전에 논문<sup>[1,2]</sup>을 통하여 알려졌지만, 이 두 이론을 실제 비디오신호에 적용한 구체적인 실험은 몇 년 뒤인 다음의 두 연구결과에 의한다. 하나는 미국의 UC Berkeley 대학의 K.Ramchandran 교수 연구팀에 의해 개발된 PRISM (Power efficient, Robust, hIgh compression, Syndrome based Multimedia coding) 부호화 방법이고<sup>[3,4]</sup>, 또 하나는 미국 스탠포드 대학의 B.Girod 교수 연구팀에 의해 개발된 Wyner-Ziv 부호화 방법이다<sup>[5,6]</sup>. 이 두 구조는 상호 상대적인 장단점을 가지고 있으나, 최근에는 대부분 Wyner-Ziv 구조를 기반으로 연구가 진행되고 있다. 그 중 대표적인 것이 2005년부터 2007년까지 수행된 유럽 연합 프로젝트 DISCOVER 프로젝트이다. 이 연구프로젝트는 Wyner-Ziv 구조를 기반으로 DVC 기술을 집중적으로 연구하여 현재 자체 코덱 프로그램을 실행코드로 공개하였다<sup>[7]</sup>. Wyner-Ziv 구조는 PRISM 구조에 비해, 상대적으로 접근하기가 보다 용이하고 관련 정보도 풍부하다. 전술한 바와 같이, 두 가지 구조는 상대적인 장단점을 가지고 있기 때문에 그 구조적인 특징을 비교하여 살펴보도록 하자.

## 2. PRISM 부호화 구조

PRISM 구조의 가장 큰 특징은 Wyner-Ziv 구조와 달리 피드백채널을 사용하지 않는 것이다. 즉, 화질, 율 등 부호화에 대한 모든 결정을 부호화기가 내린다. <그림 2>의 PRISM 코덱의 부호화기에는 일반적인 비디오 압축 부호화기에 없는 Classifier, CRC (Cyclic Redundancy Check) 생성, 신드롬(Syndrome) 부호화 부분이 있다. 또한 복호화기를 보면 복호화기 내에 움직임 추정 블록이 있는 것을 볼 수 있다.



〈그림 2〉 PRISM의 부호화기 및 복호화기 구조

PRISM은 종래 MC-DCT 하이브리드 구조의 비디오 부호화와 마찬가지로, 블록 기반 부호화 구조를 유지하고 있다. 그 이유는 부호화기에서 영상의 움직임 특성에 따라 잡음 모델링을 수행하고 이 잡음 모델을 기반으로 채널부호화를 원활하게 수행하기 위함이다. 부호화기의 Classifier는 영상의 움직임 특성을 판단하기 위해, 블록 단위로 이전 프레임의 동일 위치 블록과의 평균 제곱오차(MSE)를 계산하고 그 오차의 크기에 따라 잡음을 모델링한다. 일반적으로 영상의 움직임이 큰 영역은 배경과 같은 영역에 비해 평균 제곱오차가 크고 잡음도 이에 비례하여 커지기 때문에, 평균 오차를 이용하여 잡음의 양의 예측이 가능하다. 이러한 잡음은 DCT 계수 위치별로 모델링된다. 이때 잡음의 크기정도에 따라 인트라나 스킵(SKIP)의 부호화모드 결정도 가능하다. 이렇게 결정된 각 잡음 모델 및 부호화 모드 정보는 헤더정보로 복호화기에 전송된다.

한편, 잡음의 정도에 무관하게 일정 수준의 채널 에러를 유지하도록 하기 위하여, Classifier를 통해 모델링된 잡음의 크기에 비례하여 양자화 크기를 결정한다. PRISM구조에서의 신드롬 부호화에서는 양자화 크기가 증가함에 따라 채널

에러율도 낮아지는 특성으로 인하여 양자화 크기를 이용하여 채널 에러율의 조절이 가능하다. 잡음의 크기에 맞춰 양자화 크기를 증가시킴에 따라 부호화기에서 목표로 하는 영상의 왜곡보다 더 큰 왜곡을 초래하는 경우에는 이를 보상하기 위해 목표치 수준의 양자화 크기로 다시 보상해 주며 추가 정보가 전송된다.

이러한 잡음 모델을 기반으로 양자화된 DCT 계수의 레벨 값들은 신드롬 부호화를 통하여 coset의 인덱스가 결정되고 이렇게 결정된 인덱스값이 전송되게 된다. 양자화 크기 조절을 통해 잡음에 무관하게 채널 에러가 어느 정도 일정하게 유지되었기 때문에 신드롬 부호화에서는 이러한 채널 에러에 맞춰서 coset을 생성하고 채널 부호화를 수행한다. 또한, coset을 보다 효율적으로 전송하기 위해 Trellis 코드를 이용한 컨볼루션 부호화가 사용된다.

복호화기에서는 먼저 부호화 모드 및 잡음 모델에 대한 정보를 복호화하고 이를 기반으로 신드롬 복호화를 수행한다. Trellis 부호화된 각 coset의 인덱스를 복호화 할 때 비터비(Viterbi) 알고리즘 등 연관정(soft decision) 알고리즘을 사용하여, 보다 효율적으로 전송 에러를 극복할

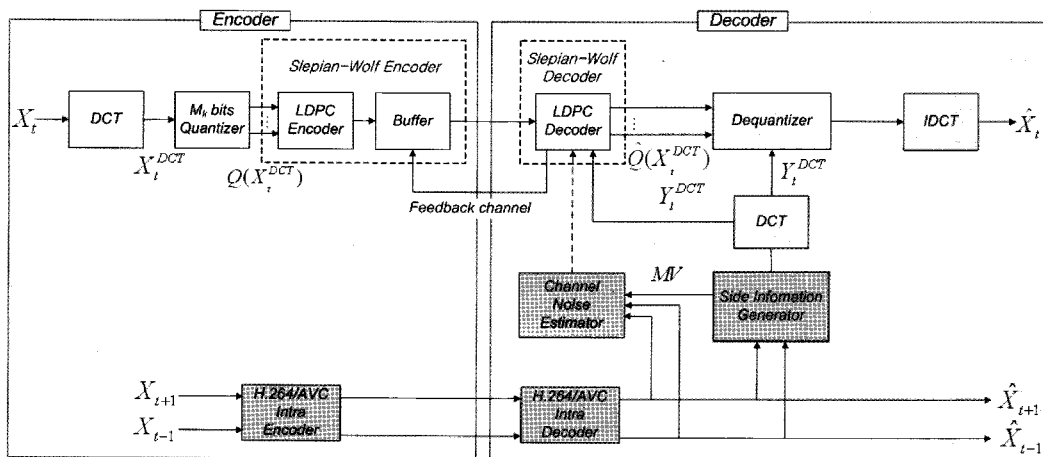
수 있다. 이렇게 각 coset의 인덱스가 복호화 된 후 coset으로부터 원래의 DCT 계수값을 찾기 위해 움직임 추정이 행해진다. 즉, 움직임 벡터 위치의 참조 블록의 DCT 계수값과 가장 오차가 작은 DCT 계수값으로 복원하게 된다. 이때 복호화는 부호화기가 전송한 CRC 값을 사용하여, 부호화기가 움직임 벡터를 전송하지 않는 데도 불구하고, 움직임 추정을 할 수 있다. 즉, 움직임 추정 및 coset을 통해 복원된 DCT 계수값이 올바르게 복원되었는지 여부를 판단하도록 원 DCT 계수값에 대해 적용한 CRC 값과 복원된 DCT 계수값에 대한 CRC 값과의 일치 여부를 확인한다. 따라서 움직임 추정 및 DCT 계수값 복원은 CRC 값이 일치할 때까지 수행된다. CRC 값 전송을 통해, 부호화기에서의 움직임 추정을 가능하게 하는 동시에 채널 부호어(coset 코드)를 통해 복원된 DCT 계수 레벨값이 부호화기에서 양자화된 원 DCT 레벨과 일치하는지의 여부를 확인할 수 있기 때문에 CRC 전송은 바람직하지는 않지만 정보의 전달과 그 비용이라는 입장에서 보면 전송율의 5-10%를 항상 차지하는 CRC

전송은 올 왜곡 성능을 저하시키는 주요 요인이기도 하다. 또한 multi-level coset 코드는 적정 에러에 대해서 강인한 특성이 있는 반면 레벨 값을 항상 알려줘야 하기 때문에 부가적으로 비트량을 증가시킨다.

### 3. Wyner-Ziv 부호화 구조

Wyner-Ziv 부호화 구조는 <그림 3>에서 볼 수 있는 것처럼 피드백채널을 가지는 것이 특징이다. 또한 Wyner-Ziv 부호화는 프레임을 두 종류로 구분하는데, <그림 3>에서와 같이,  $X_t$ 에 해당하는 프레임은 WZ (Wyner-Ziv) 프레임, 기존 부호화방식에 따라 인트라 부호화되는  $X_{t-1}$ ,  $X_{t+1}$  프레임은 Key 프레임이라고 한다.

PRISM 부호화 방법에서는, I-P-P-P의 프레임 구조에서 P 프레임에 해당하는 프레임들이 DVC로 부호화되는 반면, Wyner-Ziv 부호화 방법은 I-WZ-I-WZ 구조로 부호화하기 때문에 WZ 프레임의 앞과 뒤에 항상 인트라 부호화



<그림 3> 변환 영역 Wyner-Ziv 부호화기 및 복호화기 구조

된 Key 프레임이 필요하다. Key 프레임  $X_{t-1}$ 과  $X_{t+1}$ 은 H.264/AVC 등의 기존 인트라 부호화 방식으로 부호화 되어 복호화기로 전송되며, 복호화기는 이 두 장의 프레임을 움직임 추정 보간법 (Motion Compensated Temporal Interpolation, MCTI 혹은 MCI라고 함)을 이용하여 현재 프레임  $X_t$ 에 상응하는 보조정보인 Y 프레임을 만들어 낸다.

<그림 3>에서 보는 바와 같이, 부호화기는 DCT, 양자화, 채널 부호화의 3가지 부호화 과정을 수행하기 때문에 상당히 간단하며, PRISM 구조에서 볼 수 있었던 Classifier와 같은 판단/제어 모듈이 없기 때문에 실제 잡음의 양 혹은 잡음 모델의 추정 등을 복호화기가 수행한다. 따라서 Wyner-Ziv 구조는 복호화기 주관 부호화를 수행한다고 이해할 수 있다. 또한 PRISM과는 다르게 블록 기반의 부호화가 아닌 프레임 기반 부호화를 하고 있는 것도 또 다른 특징으로 볼 수 있는데, 이것은 Turbo 코드나 LDPC 등의 고성능 코드를 사용하여 효율적 채널 부호화를 하기 위함이다.

Wyner-Ziv 구조에 따른 부호화 동작을 살펴보면, 앞서 언급한 바와 같이  $X_{t-1}$ 과  $X_{t+1}$ 은 인트라로 부호화되어 복호화기로 전송되고 현재 프레임인  $X_t$ 는 DCT 변환과 M비트 양자화를 거치고 주파수 밴드별로 DCT 계수가 모아진 뒤 각 밴드의 비트플레인 별로 패리티 데이터가 생성된다. QCIF 영상(176x144)의 경우를 예를 들어 설명하면, 4x4 DCT 변환을 통해 16개의 주파수 밴드를 생성하고, 각 밴드를 양자화한 결과가 계산의 편의상 4비트라고 가정하면, 최종적으로 생성되는 패리티의 길이는 각 주파수 밴드에 대해 각각 1584 비트의 코드가 4개씩 만들어진다. 이렇게 채널부호화가 끝나면 버퍼에 비트스

트림을 저장하고 복호화기로 부터의 송신요청 신호를 기다린다.

복호화기는 전송받은  $X_{t-1}$ 과  $X_{t+1}$ 을 인트라 복호화기를 통해 복호화하고, MCTI를 통해  $X_t$ 에 해당하는 보조 정보 Y를 생성한 후 DCT 변환한다. 이후 복호화기는 피드백채널을 통하여 부호화기에 부분적으로 채널코드를 요청하는데, 채널 복호화를 통해 얻어진 에러율이 충분히 작아질 때까지 부호화기에 패리티를 계속 추가로 요청한다. 앞서 설명했던 것처럼 패리티의 길이는 1584 비트인데 반해 부분적으로 패리티를 요청하고 채널 복호화에 이용할 수 있는 이유는 현재 사용하고 있는 Turbo 코드의 경우 RCPT (Rate Compatible Punctured Turbo) 코드를 사용하고 있고, LDPC (Low Density Parity Check) 코드의 경우에는 LDPCA 코드를 사용하고 있기 때문에 패리티를 부분적으로 전송받아 채널 복호화를 수행하는 것이 가능하기 때문이다. LDPCA 코드를 예를 들어 설명하면, 현재 24비트의 패리티 블록이 66개가 있기 때문에 한 주파수 밴드에 대한 한 비트플레인의 경우 최대 66회의 패리티 요청이 가능하다.

현재 채널 복호화 과정 중에 반드시 필요한 정보는 채널 잡음 모델에 대한 것이다. 채널 복호화를 하기 위해 필요한 요소는 패리티, 원본 정보에 잡음이 추가된 것으로 가정하는 보조 정보, 그리고 채널 잡음 모델, 이렇게 세 가지이다. 하지만 현재와 같은 Wyner-Ziv 구조에서는 실제 잡음 모델을 알 수가 없는데, 그 이유는 실제 잡음 모델은 원본 정보인 X와 보조 정보 Y의 차이를 통해서만 알 수 있기 때문이다. 다시 말하면 보조 정보 Y는 복호화기에서 만들어지지만 원본 정보인 X는 부호화기에서만 알 수 있기 때문에 정확한 잡음 모델  $P(X-Y)$ 는 부호화기와 복호화기

가 양쪽 다 알 수 없다. 따라서 현재 구조에서는, 채널 잡음 모델을 추정할 수 밖에 없으며, 이를 위한 가장 간단한 방법은 앞과 뒤의 Key 프레임의 차이인  $P\{X_{t-1}-X_{t+1}\}$ 를  $P\{X-Y\}$ 로 대신하여 이용하는 것이다.

채널 복호화가 끝나면, 양자화 된 계수들을 채널 잡음 모델을 이용하여 최적으로 복원한다.

Wyner-Ziv 구조에서의 복호화기는 기존 부호화 방식에 비해 복잡도가 증가하는 것이 일반적인데, 그 이유는 크게 두 가지로 나눌 수 있다. 첫 번째는 MCTI 과정을 들 수 있는데, MCTI 과정은 실제로 양방향 움직임 추정 과정과 상당히 유사하며 반복적으로 수행되는 특징을 갖기 때문에 복잡도를 증가시키는 주요 요인이 된다. 다른 한 가지는 채널 복호화과정을 들 수 있으며 이는 현재 사용하고 있는 Turbo 코드나 LDPC 코드의 경우 Shannon 바운드에 근접한 좋은 성능을 내는 반면, 그 복잡도가 상당히 증가하기 때문이다. 또한 앞서 언급한 것과 같이 이러한 Shannon 바운드 근접 채널 코드는 입력신호의 길이가 짧을 경우에는 그 성능이 저하되는 특징으로 인해 블록 단위(짧은 길이)의 부호화가 불가능하기 때문에 Wyner-Ziv 구조는 프레임 기반 부호화 방식으로 불가피하게 바뀌었다고 볼 수 있다.

Wyner-Ziv구조와 PRISM 구조에 대한 또 다른 상대적인 차이점은 그 대상 응용이다. 즉, 성능과 관계없이 그 구조 자체만을 놓고 봤을 때 PRISM 구조는 오프라인 응용에 적합하다 할 수 있고, Wyner-Ziv 구조는 온라인 응용에 적합하다고 할 수 있다. <표 1>은 앞서 언급했던 PRISM 구조와 Wyner-Ziv 구조의 상대적인 특징을 항목별로 요약한 것이다.

<표 1> PRISM과 Wyner-Ziv 구조의 특징

PRISM	Wyner-Ziv
I-P-P-P	I-WZ-I-WZ
피드백채널 없음	피드백채널 있음
블록 단위 부호화 (국부적 적응성)	프레임 단위 부호화 (전체적 적응성)
INTRA/SKIP 가능	INTRA/SKIP 적용이 어려움
부호화기 주관	복호화기 주관
offline 응용 가능	online 응용만 가능

### III. Wyner-Ziv구조에 대한 주요 연구동향

본 장에서는 상대적으로 많은 연구가 이루어지고 있는 Wyner-Ziv 구조를 중심으로 부호화 성능에 영향을 미치는 부/복호화 과정을 부분별로 나누어 현재까지의 연구결과를 소개한다.

#### 1. 양자화

Wyner-Ziv 구조에서는 균일 양자화기를 사용한다. 4x4 DCT 변환을 통해서 16개의 밴드가 생성되는데 이중 DC 계수의 경우에는 계수의 범위를  $0 \sim 2^{11}-1$ 까지이며, AC 계수의 경우에는  $-Max_b \sim +Max_b-1$  까지 계수의 범위이다. 또한 AC의 경우에는 데드존(dead-zone) 양자화기를 적용하며 데드존의 범위는 양자화 스텝 사이즈의 2배로 한다.

[DC 양자화]

l 번째 양자화 구간:  $I_{DC} = [l\Delta_{DC}, (l+1)\Delta_{DC})$

DC 양자화 스텝 사이즈:  $\Delta_{DC} = 2^{(11-M_6)}$

16	8	0	0	32	8	0	0	32	8	4	0	32	16	8	4
8	0	0	0	8	0	0	0	8	4	0	0	16	8	4	0
0	0	0	0	0	0	0	0	4	0	0	0	8	4	0	0
0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
(1)				(2)				(3)				(4)			
32	16	8	4	64	16	8	8	64	32	16	8	128	64	32	16
16	8	4	4	16	8	8	4	32	16	8	4	64	32	16	8
8	4	4	0	8	8	4	4	16	8	4	4	32	16	8	4
4	4	0	0	8	4	4	0	8	4	4	0	16	8	4	0
(5)				(6)				(7)				(8)			

〈그림 4〉 Wyner-Ziv 구조에서 사용하는 양자화 매트릭스<sup>[23]</sup>

[AC 양자화]

l 번째 양자화 구간 :

$$I_b^l = \begin{cases} [(l-1)\Delta_b, l\Delta_b), & l < 0 \\ [-\Delta_b, \Delta_b), & l = 0 \\ [l\Delta_b, (l+1)\Delta_b), & l > 0 \end{cases}$$

AC 양자화 스텝 사이즈:  $\Delta_b = \lceil \frac{2Max_b}{2^{M_b}} \rceil$

여기서  $M_b$ 는 양자화 후에 남은 비트수를 의미하며,  $b$ 는 DC에서 AC에 이르는 밴드의 인덱스이다. 또한 각 밴드별로 양자화 스텝 사이즈는 이미 정해져 있는데, 저비트율에서 고비트율에 이르기까지 <그림 4>의 (1)에서 (8)까지가 사용된다. 예를 들어 (8)에 보여지는 양자화 매트릭스에서 DC의 경우를 보면, DC 계수에 해당하는 128은 양자화 후에 남겨지는 양자화 계수가 0~127까지임을 의미하며 따라서 실제 양자화 스텝 사이즈는  $2^{(11-7)} = 16$  이 된다. 동일하게 AC의 마지막 0 값은 양자화 후에 남겨지는 값이 없다는 것을 의미하므로 <그림 4>에 있는 8개의 양자화 매트릭스에서 0 값을 가지는 위치의 DCT 밴드는 양자화 후에 아무것도 전송되지 않는다. 즉, 채널부호화가 수행되지 않는 밴드이다.

## 2. 보조 정보(Side Information) 생성

보조 정보 생성 과정은 일반 비디오 압축 기술에서의 움직임 추정과 그 결과 생성되는 움직임 보상 신호를 생성하는 과정과 유사하다. 정확한 예측자(predictor)를 찾아내면 잔차 오류(residual error) 신호의 양이 줄어드는 것과 유사하게 좋은 보조 정보를 생성하면 부호화기에서 전송하는 패리티의 양이 줄어든다. 따라서 보조 정보 생성 과정은 분산 비디오 압축 기술의 성능에 가장 직접적으로 영향을 미치는 부분이다. 보조 정보 생성에 대해서는 상당히 많은 연구 결과가 있는데<sup>[8-11]</sup>, 아직은 절대적으로 크게 뛰어난 결과를 보이는 논문은 많지 않다. 다시 말하면 보조 정보 생성이 그 만큼 어려운 문제라는 것을 단적으로 보여주는 것이라고 할 수 있다.

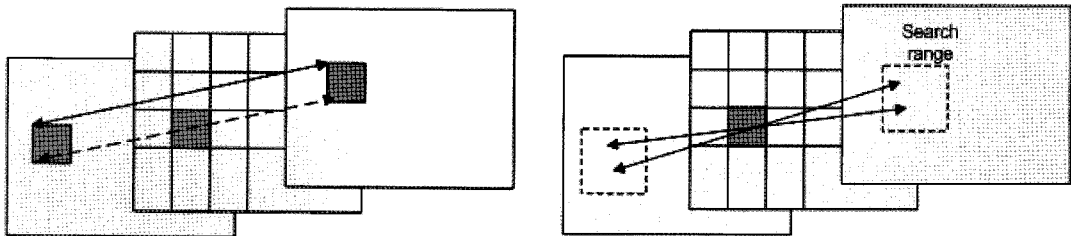
보조 정보 생성문제는 일반 비디오 부호화 과정에서 움직임 추정과 비교해 보면 그 주요 문제점을 파악하기 용이하다. 일반적인 움직임 추정의 문제는 원본 프레임 X에 속하는 하나의 블록 Bi에 대해 참조 영상  $X_{t-1}$ 에 속하는 후보 블록

인  $C_1, C_2, \dots, C_N$  이 주어질 때,  $B_i$ 와 비교하여 가장 왜곡이 적은 후보 블록을 고르는 문제인 반면, 보조 정보 생성의 문제는 원 프레임  $X_t$  가 주어지지 않은 상황에서 이전 키 프레임  $X_{t-1}$  과 이후 키 프레임  $X_{t+1}$  을 이용하여  $X_t$  와 가장 유사한 보조 정보  $Y_t$  를 생성하는 문제이다. 즉,  $X_t$  와 가장 유사한(왜곡이 적은)  $Y_t$  를 생성해야 함에도 불구하고 실제 왜곡의 정도를 측정할 기준이 되는  $X_t$  가 주어지지 않기 때문에 어려운 문제이다. 이와 유사한 문제로 프레임 레이트 업 변환(Frame rate up conversion)이 있지만 프레임 레이트 업 변환의 경우에는 시각적인 오류를 줄여야 하는 반면 보조 정보 생성의 경우는 원 프레임  $X_t$  와 얼마나 유사하게 만드는가의 문제가 기 때문에 좀 더 높은 정확도를 요구한다.

보조 정보 생성 과정은 전방향 움직임 추정과 양방향 움직임 추정, 그리고 공간적 움직임 평활화(smoothing) 과정을 거친 후 최종적으로 양방향 움직임 보상이 수행된다. 전방향 움직임 추정 과정에 앞서 저역 통과 필터(Low Pass Filter)를 취하는데 움직임 벡터 필드의 오류를 줄여주는 역할을 한다. <그림 5>(a)와 같이 이전 프레임  $X_{t-1}$  과 이후 프레임  $X_{t+1}$  간의 움직임 추정을 하게 되면 현재 프레임  $X_t$  의 각 블록  $B_i$  를 통과하는 움직임 벡터가 생기게 되는데 이를 블

록  $B_i$ 의 움직임 벡터로 간주한다. 한편, 한 블록 내에 움직임 벡터가 전혀 통과하지 않는 경우도 발생하는데, 이 경우에는 현재 블록의 중앙에서 가장 가까운 움직임 벡터를 현재 블록의 움직임 벡터로 간주한다. 이 과정이 끝나면 현재 프레임  $X_t$ 의 모든 블록에는 움직임 벡터가 하나씩 생기게 되며 양방향 움직임 추정을 통해서 현재의 움직임 벡터를 보다 정교하게 만든다. 양방향 움직임 추정 시에는 <그림 5>(b)의 예처럼 현재 프레임의 블록의 중앙에 움직임 벡터의 중간을 고정시키고, 이전 및 이후 프레임에 움직임 벡터가 항상 대칭되게 움직이도록 한다. 다음 과정은 움직임 평활화 과정인데 가중 벡터 미디언 필터(weighted vector median filter)를 적용한다. 가중치는 블록 정합 시에 계산되는 MSE 값을 이용하여 현재 블록과 주변 블록의 각각의 MSE 값의 비(ratio)로 구한다. 움직임 평활화 과정이 끝나면 결정된 움직임 벡터를 이용하여 양방향 움직임 보상 과정을 수행하여 최종적인 보조 정보를 생성한다.

보조 정보 생성시에는 현재 프레임이 없는 상태에서 움직임 추정을 하기 때문에 프레임간의 움직임은 선형적이라는 가정을 한다. 따라서 선형적인 영상의 경우에는 높은 화질의 보조 정보가 생성 되지만 비선형적인 움직임이 있거나 겹



(a) 전방향 움직임 추정

(b) 양방향 움직임 추정

<그림 5> 전방향 및 양방향 움직임 추정



침(occlusion)이 발생하는 경우는 이러한 가정이 맞지 않는다. 특히 복잡한 움직임은 가지는 영상의 경우는 각 프레임마다의 움직임의 선형성 없거나 여러 방향의 움직임이 프레임 전체에 걸쳐 이질적으로 생기기 때문에 이런 경우는 아주 낮은 화질의 보조 정보가 생성된다. 이러한 문제를 해결하기 위해 해쉬(Hash) 정보를 부호화기에서 전송하는 방법으로<sup>[9]</sup> 보조 정보의 화질을 상당히 개선시킬 수 있다. 하지만 실제 해쉬가 필요한 블록과 필요 없는 블록, 그리고 적절한 해쉬의 양의 결정이 부정확한 경우 전송 비트량을 증가시키기 때문에 전체적인 유휴 용량의 관점에서 보면 개선 효과는 크지 않을 수 있다. 이와 유사한 연구로는 현재 프레임을 낮은 화질로 부호화(Low Quality Reference)하여 전송한 후 복호화기가 보조 정보를 생성할 때 이를 이용하는 방법<sup>[10]</sup>도 있다. 이 외에도 채널 복호화된 정보를 이용하여 보조 정보를 개선하는 방법도 있는데, 그 구체적인 예로, 복호화 시에 점진적 정교화(successive refinement) 방법<sup>[11]</sup>이나 EM(Expectation Maximization) 알고리즘을 이용하여 보조 정보를 개선하는 방법<sup>[12]</sup> 등이 있다.

### 3. 채널 잡음 모델 추정

채널 복호화 시에는 보조 정보와 채널 잡음 모델을 이용하여 소프트 입력(soft input) 값을 계산하고 이를 기반으로 채널 복호화기가 복호화를 수행하게 된다. 따라서 정확한 채널 잡음 모델 추정은 보조 정보의 생성 다음으로 분산 비디오 압축 기술에서 중요한 부분이다. 채널 잡음  $N$ 은 원본 영상을  $X$ , 보조 정보를  $Y$ 라고 할 때  $Y=X+N$ 의 관계를 가지며 통신 채널에서 일반적

으로 생각할 수 있는 실제 잡음이 아닌 원본과 보조 정보 사이에 가상적으로 추가된 잡음이기 때문에 가상 채널 잡음(virtual channel noise)이라고도 한다. 채널 잡음 모델로는 라플라시안(Laplacian) 분포를 사용한다.

채널 잡음 모델은 역시 복호화기에서는  $X$ 를 알 수 없고 부호화기에서는  $Y$ 를 알 수 없기 때문에 추정치를 사용할 수밖에 없는데, 현재 사용되는 방법은 복호화기에 전송된 이전과 이후 프레임인  $X_{t-1}$ 과  $X_{t+1}$ 의 차이 값을 이용한다. 즉  $X_{t-1}$ 과  $X_{t+1}$ 로부터 움직임 보상된 프레임을 각각  $Y_b$ 와  $Y_f$ 라 하면  $P\{X-Y\}$  대신  $P\{Y_b-Y_f\}$ 를 추정된 채널 잡음 모델로 사용하는 것이다.  $P\{X-Y\} = P\{Y_b-Y_f\}$ 의 관계는 두 가지 가정 아래에서는 정확성을 가지는데, 첫 번째 가정은 보조 정보 생성 시와 마찬가지로 시간적인 움직임의 선형성이며 두 번째 가정은 화소값 변화의 시간적인 선형성이다. 따라서 이 두 가지 조건(가정)을 만족하는 한, 채널 잡음 모델은 높은 정확도를 유지한다.

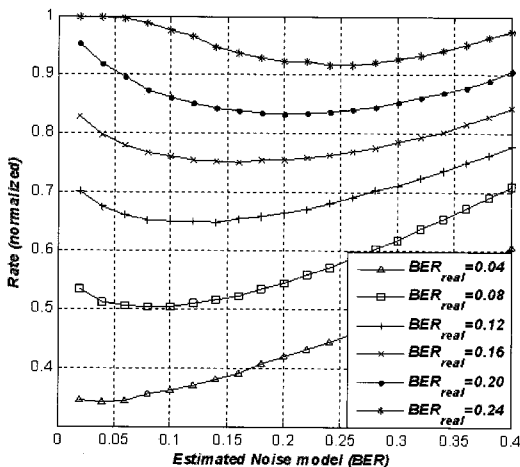
채널 잡음 모델인 라플라시안 분포는  $Y_b$ 와  $Y_f$ 의 차이에 대해 각 DCT 밴드별로 파라미터  $\alpha$  값을 구해야 하며 이것은 밴드별 분산을 통해서  $\alpha_b = \sqrt{2}/\sigma_b$ 와 같이 구할 수 있다. 여기서 첨자  $b$ 는 각 밴드를 의미한다. 따라서 밴드별로 각각 하나의 라플라시안 분포가 결정되는데 이렇게 밴드 전체에 대해서 채널 잡음 모델을 사용하는 것에 비해 밴드의 공간적인 변화에 따라 적응적인 잡음 모델을 사용하는 것이 효과적이기 때문에 최종적으로 사용되는 채널 잡음 모델의 라플라시안 파라미터는 다음과 같이 결정한다<sup>[13]</sup>.

$$\alpha_b(u,v) = \begin{cases} \alpha_b & , [D_b(u,v)]^2 \leq \sigma_b^2 \\ \sqrt{2/[D_b(u,v)]^2} & , [D_b(u,v)]^2 > \sigma_b^2 \end{cases}$$

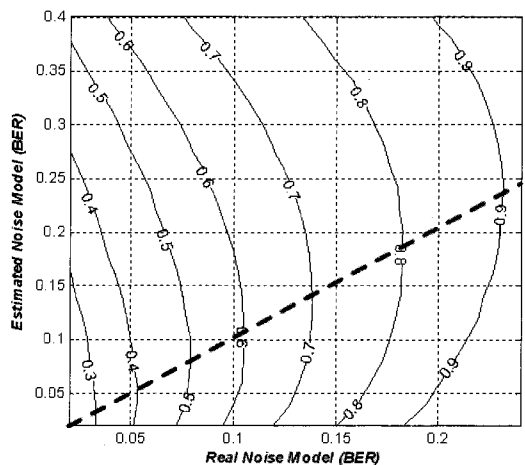
여기서  $\alpha_b(u,v)$ 는 DCT 계수의  $(u,v)$  위치에 서의 라플라시안 파라미터를 의미하며,  $D_b(u,v)$ 는 DCT 밴드  $b$ 의 계수들의 평균값과  $(u,v)$  위치의 계수 값과의 차이를 말한다.

앞서 언급했던 것처럼 채널 잡음 모델은 움직임과 화소 값 변화의 시간적 선형성이라는 두 가지 가정을 만족하면 실제 채널 잡음 모델과 유사해지지만, 실질적으로 움직임이 선형적이 아닌 경우가 많고, 화소 값이 비선형적으로 변하는 경우가 더 많다. 따라서 부정확하게 추정된 채널 잡음 모델을 이용하여 채널 복호화를 수행하는 경우가 많다. <그림 6>은 추정된 채널 잡음 모델이 실제 채널 잡음 모델과 차이가 날 경우, 채널 복호화 성능에 얼마나 영향을 미치는 지 보여주는 그래프이다. 두 그래프 모두 LDPCA를 사용해서 복호화한 경우이다. <그림 6>의 (a) 그래프에서 x축은 채널 복호화 시에 입력으로 주어진 채널 잡음 모델의 BER 값이며 y축은 오류 정정을 위해 요구된 패리티의 양을 정규화한 값이다. 또한 6개의 선은 실제 채널 잡음 모델의 BER 값이 각

각 0.04에서 0.24까지(아래에서 위의 순서) 0.04의 간격으로 생성한 보조 정보를 의미한다. 맨 아래의 선의 예를 들면 실제 잡음 모델에 해당하는 BER이 0.04인 경우이며, 실제 BER이 아닌 x축에 따른 BER을 채널 복호화 시 입력으로 주었을 때 늘어나는 패리티의 양이다. 따라서 실제 잡음인 BER = 0.04로 채널 복호화할 경우 필요한 패리티 양이 0.34정도로 가장 낮으며, BER = 0.04 대신 BER = 0.24로 채널 복호화할 경우 패리티의 양이 0.45로 대략 30% 정도 늘어나는 것을 볼 수 있다. <그림 6>의 (b) 그래프는 (a)와 동일하지만 다른 각도에서 본 그래프이다. x축은 실제 잡음 모델이며, y축은 입력으로 넣어주는 추정 잡음 모델이다. 여기서의 각각의 선은 동일한 패리티의 양을 필요로 하는 지점들을 연결할 것이며, 점선은 실제 잡음 모델과 추정 잡음 모델이 동일한 지점을 연결한 선이다. 따라서 맨 오른쪽에 위치한 선의 경우 그 선이 지나가는 모든 지점에서는 채널 복호화 시 90%의 패리티 양이 필요한 것이다. 따라서 실제 BER이



(a) 추정잡음모델에 따른 패리티량의 변화



(b) 실제 및 추정잡음모델에 따른 패리티량의 변화

<그림 6> 추정 잡음 모델과 실제 잡음 모델에 따른 패리티 양의 변화

0.16인 보조 정보라도 정확한 잡음 모델이 주어지지 않으면 실제 BER이 0.24인 경우(잡음이 50% 증가한 경우)와 복호화에 필요한 패리티의 양이 동일해 질 수 있다는 것이며, 추정 잡음 모델의 정확도에 따라 윌 성능이 얼마나 저하될 수 있는지를 보여주는 예이다.

#### 4. 모드 결정 및 기타 이슈

앞에서 다룬 Wyner-Ziv 구조에서의 주요 연구 방향 외에도 다른 중요한 연구 주제나 방향에 관한 연구들도 찾아 볼 수 있는데, 그 중 한 가지는 모드 결정에 관한 문제이다. 앞서 언급한 Wyner-Ziv의 특징 중에서 프레임 단위의 부호화가 이루어지기 때문에 전체적 적응성(혹은 프레임 단위의 적응성)을 가진다고 요약했지만, 이것은 PRISM 구조와 비교했을 때 상대적인 특징이며, 이는 국부적 적응성(혹은 블록 단위 적응성)이 미비함을 의미한다. 따라서 Wyner-Ziv 구조에서는 모드별 부호화 구조를 구현함에 있어 어려움이 있으며, 이는 보조 정보를 생성하거나 채널 잡음 모델을 추정할 때와 동일하게 현재 Wyner-Ziv 구조 자체로 인해 발생하는 문제이다. 다시 말하면 블록 단위로 적응적인 부호화를 수행하기 위해서는 예측 에러를 보다 정확히 파악할 수 있어야 하지만, Wyner-Ziv 부호화기는 보조 정보 Y에 대한 정보를 가지고 있지 않기 때문에 효과적인 모드 결정이 어렵다.

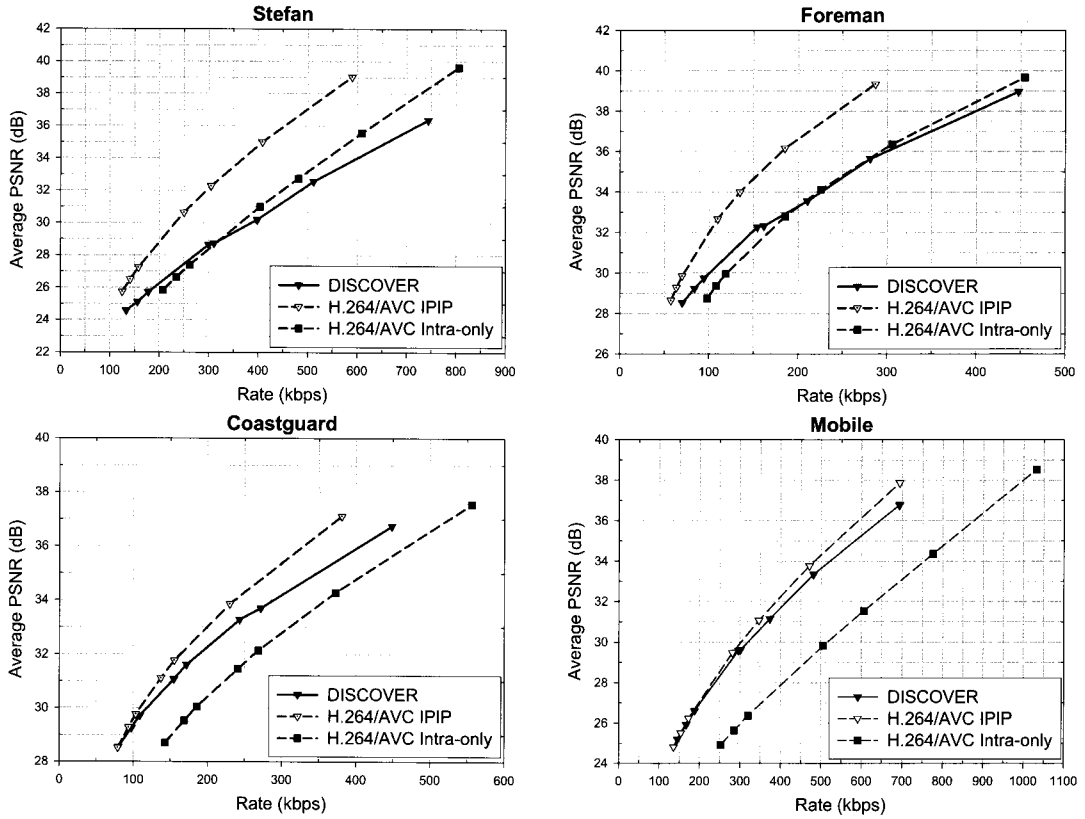
Wyner-Ziv에 적용이 가능한 또는 필요한 모드는 크게 인트라 모드와 스킵(SKIP) 모드를 들 수 있다. 인트라 모드는 윌 성능과 왜곡 성능에 동시에 영향을 미치게 된다. 이는 분산 비디오 압축 기술이 엔트로피 부호화 대신 채널 부호화를 사용하기 때문에, 채널 용량을 넘어서는 예

러가 발생할 경우 패리티의 증가뿐만 아니라 복구할 수 없는 화질의 열화까지 발생하기 때문이다. 스킵 모드는 일반적인 비디오 부호화의 경우와 동일한 이유로 윌 성능 면에 있어서 중요하다.

이러한 모드 결정에 관한 연구들은 부호화기에서 이전 프레임의 동일 위치 블록과의 비교를 통해서 선택하는 방법<sup>[14]</sup>, 영상의 움직임에 기반하여 프레임 자체를 인트라 코딩하는 방법<sup>[15]</sup>, 이전 프레임과의 잔차 신호(residual)에 기반한 분산 비디오 압축 기술에서의 스킵 모드 결정 방법<sup>[16]</sup>, 윌 추정 모델을 기반으로 반복 알고리즘(iterative algorithm)에 의한 모드 결정 방법<sup>[17]</sup> 등 다양하게 수행되었다.

모드 결정 방법 이외에도 일반적인 비디오 부호화에서 사용되는 다양한 알고리즘들을 분산 비디오 압축 기술에 적용한 연구 결과도 찾아볼 수 있는데, 보조 정보 생성 시에 가변크기블록을 이용하여 화질을 향상시키는 방법<sup>[18]</sup>, 보조 정보 생성 시에 다중 참조 영상(multiple reference)를 이용하여 화질을 향상시키는 방법<sup>[19,20]</sup>, GOP 단위의 변화에 따른 성능에 대한 연구, 부호소 정확도(sub-pel precision)에 따른 보조 정보 생성에 대한 연구<sup>[19]</sup> 등 다양하다. 이러한 연구들의 공통적인 특징은 기존 MC-DCT 기반 압축 기술의 경우 부호화기에서 고려하던 여러 가지 알고리즘을 Wyner-Ziv 구조에서는 복호화기에 어떻게 적용할 것인가에 대한 해답을 찾고 있다는 것이다.

<그림 7>은 2007년 종료된 DISCOVER 프로젝트의 윌-왜곡 성능을 보여준다. DISCOVER 프로젝트 이후에도 많은 연구 결과들이 발표되었지만 DISCOVER-DVC (Distributed Video Coding : 분산 비디오 압축 기술) 코덱<sup>[7,21,22]</sup>은 여전히 현재 Wyner-Ziv 구조에서의 성능을 단



〈그림 7〉 DISCOVER-DVC의 올 왜곡 성능

적으로 보여주는 좋은 예이다. <그림 7>과 같이, DISCOVER-DVC 성능을 살펴보면 영상에 따라 편차가 심한 것을 볼 수 있는데, 이는 현재의 분산 비디오 압축 기술의 성능이 영상의 특성에 따라 좌우되고 있음을 잘 나타내고 있다. 예를 들어, 선형적인 특성이 강한 순서로 영상을 나열해 보면 Mobile, Coastguard, Foreman, Stefan의 순서라고 할 수 있는데, DISCOVER DVC의 성능도 동일한 순서로 우수한 성능을 보여준다 (인트라와 I-P-I-P 대비 상대적으로). 특히, Foreman 이나 Stefan 영상의 경우를 살펴보면, H.264/AVC 인트라 수준이거나 또는 인트라에 못 미치는 성능을 보이고 있다. 이는 분산 압축 기술 분야에 있어 성능 향상 문제가 여전히 존재

하고 있음을 의미한다. 따라서 향후의 분산 비디오 압축 기술 분야에서 연구가 진행될 방향은 주로 성능 개선의 방향이라 할 수 있다. 또한 현재의 Wyner-Ziv 구조는 구조적 특성상 feedback 채널을 항상 유지해야 한다. 뿐만 아니라, 현재까지 개발된 분산 비디오 압축 기술의 복잡도가 부호화기의 경우에는 H.264/AVC 인트라에 비해 70% 수준으로 경량화되어 있지만, 복호화기의 경우는 H.264/AVC 복호화기에 비해 수백 배의 수준이다. 따라서 이러한 비현실적인 구조적인 한계(오프라인 응용에 제한적, 복호화기의 극심한 복잡도)에 기인하는 실제적 적용성 문제 또한 연구가 동반되어야 할 부분이다.



#### IV. 맺음말

1970년대의 Slepian-Wolf 이론을 바탕으로 30년 뒤인 2000년대에 처음으로 실제 응용을 위한 연구가 UC Berkeley 대학에서 시작되었으며, 그 뒤 몇 년 뒤 Stanford 대학을 중심으로 연구가 되었다. 하지만 많은 관심을 끌며 우수 연구 기관에서 본격적인 연구를 촉발시킨 계기는 2005년 DISCOVER 프로젝트의 시작이라고 할 수 있다. 본 고에서는 DISCOVER 프로젝트의 주요 연구주제에 대하여 양자화기, 보조 정보 생성 방법, 채널 잡음 모델 추정 방법, 그리고 모드 결정에 관한 문제 등을 위주로 현재까지 개발되어 있는 연구 결과와 드러난 문제점 등을 서술하였다. 현재 분산 비디오 압축 기술은 울 왜곡 성능을 통해서 볼 때 아직 성능 개선의 여지가 많이 남아 있다는 것을 알 수 있다. 또한, feedback 채널이 항상 필요하기 때문에 오프라인 응용에 제한적이며 따라서 구조적인 개선에 대한 부분도 앞으로 필요한 연구의 방향이라고 할 수 있다. 또한 기술의 단계적 분류 상 연구 초기 단계에 해당하기 때문에 여전히 원천기술에 대한 선점이 가능므로 이 분야에 대하여 앞으로 더 활발한 연구가 필요하다.

#### 참고문헌

- [1] J. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources," IEEE Trans. on Information Theory, Vol.19, No.4, pp.471-480, July, 1973.
- [2] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," IEEE Trans. on Information Theory, Vol.22, No.1, pp.1-10, January, 1976.
- [3] A. Aaron, R. Zhang and B. Girod, "Wyner-Ziv Coding of Motion Video," Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, November, 2002.
- [4] B. Girod, A. Aaron, S. Rane and D. Rebollo Monedero, "Distributed Video Coding," Proceedings of the IEEE, Vol.93, No.1, pp.71-83, January, 2005.
- [5] R. Puri and K. Ramchandran, "PRISM: A New Robust Video Coding Architecture Based on Distributed Compression Principles," 40th Allerton Conference on Communication, Control and Computing, Allerton, USA, October 2002.
- [6] R. Puri, A. Majumdar and K. Ramchandran, "PRISM: A Video Coding Paradigm with Motion Estimation at the Decoder," IEEE Trans. on Image Process., Vol.16, No.10, pp.2436-2448, Oct., 2007.
- [7] DISCOVER Webpage : <http://www.img.lx.it.pt/~discover/home.html>
- [8] J. Ascenso, C. Brites and F. Pereira, "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding," 5th EURASIP Conf. on Speech and Image Processing, Multimedia Communications and Services, Smolenice, Slovak Republic, June, 2005.
- [9] A. Aaron and B. Girod, "Wyner.Ziv video coding with low-encoder complexity,"

- Picture Coding Symposium, San Francisco, CA, USA, December 2004.
- [10] E. Martinian, A. Vetro, J.S. Yedidia, J. Ascenso, A. Khisti and D. Malioutov, "Hybrid distributed video coding using SCA codes," IEEE Workshop on Multimedia Signal Processing, Victoria, BC, Canada, October, 2006.
- [11] J. Ascenso, C. Brites and F. Pereira, "Motion Compensated Refinement for Low Complexity Pixel Based Distributed Video Coding," IEEE Intl. Conf. on Advanced Video and Signal Based Surveillance, Sardinia, Italy, September, 2005.
- [12] D. Varodayan, D. Chen, M. Flierl and B. Girod, "Wyner-Ziv coding of video with unsupervised motion vector learning," EURASIP Signal Processing Image Communication, Vol.23, No.5, pp.369-378, June, 2008.
- [13] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding," IEEE Trans. Circuits and Systems for Video Tech., Vol.18, No.9, pp.1177-1190, September, 2008.
- [14] G. Hua, and C. W. Chen, "Distributed Video Coding with Zero Motion Skip and Efficient DCT Coefficient Encoding," IEEE Intl. Conf. Multimedia Expo, Hannover, Germany, June, 2008.
- [15] J. Ascenso, C. Brites and F. Pereira "Content Adaptive Wyner-Ziv Video Coding Driven by Motion Activity", IEEE International Conference on Image Processing, Atlanta, USA, October 2006.
- [16] S. Mys, J. Slowacka, J. Skorupaa, P. Lamberta and R. V. Wallea, "Introducing skip mode in distributed video coding," Signal Processing Image Communication, Vol.24, Issue 3, pp.200-213, March, 2009.
- [17] L. Liu, D. He, A. Jagmohan, L. Lu, and E. J. Delp, "A Low Complexity Iterative Mode Selection Algorithm for Wyner-Ziv Video Compression," IEEE ICIP, October, 2008.
- [18] S. Argyropoulos, N. Thomos, N. V. Boulgouris and M.G.Strintzis, "Adaptive Frame Interpolation for Wyner-Ziv Video Coding," IEEE 9th Workshop on MMSP, Chania, Crete, Greece, October, 2007.
- [19] Z. Li, L. Liu, and E. J. Delp, "Rate distortion analysis of motion side estimation in Wyner-Ziv video coding," IEEE Trans. on Image Processing, Vol.16, pp.98-113, January, 2007.
- [20] A.B.B. Adikari, W.A.C. Fernando and W.A.R.J. Weerakkody, "Multiple Side Information Streams for Distributed Video Coding," IEE Electronics Letters, Vol.42, Issue 25, December, 2006.
- [21] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov and M. Ouaret, "The DISCOVER codec: Architecture, Techniques and Evaluation," Picture Coding Symposium, Lisbon, Portugal, November, 2007.

- [22] F. Pereira, C. Brites, J. Ascenso and M. Tagliasacchi, "Wyner-Ziv video coding: A review of the early architectures and further developments," IEEE Intl. Conf. on Multimedia and Expo, Hannover, Germany, June, 2008.
- [23] C. Brites, J. Ascenso and F. Pereira, "Improving transform domain Wyner-Ziv coding performance," IEEE International Conf. on Acoustics, Speech and Signal Processing, Toulouse, France, May, 2006.

저자소개



심 혁 재

2000년 2월 성균관대학교 전자공학과 학사  
 2002년 2월 성균관대학교 정보통신공학부 석사  
 2002년 3월 ~ 현재 성균관대학교 정보통신공학부 박사  
 과정

주관심 분야 : 멀티미디어 영상압축, 신호처리



전 병 우

1985년 2월 서울대학교 전자공학과 학사  
 1987년 2월 서울대학교 전자공학과 석사  
 1992년 2월 Purdue Univ, School of Elec. 공학박사  
 1993년 7월 ~ 1997년 8월 삼성전자 신호처리연구소 수  
 석연구원

1997년 9월 ~ 현재 성균관대학교 정보통신공학부 교수

주관심 분야 : 멀티미디어 영상압축, 영상인식, 신호처리