



저작자표시-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.
- 이 저작물을 영리 목적으로 이용할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

모바일 영상통화 환경에 적합한 하이브리드
변환을 이용한 객체 기반 비디오 코덱 설계



2010년 2월

부 경 대 학 교 대 학 원

컴 퓨 터 공 학 과

전 성 혜

공학석사 학위논문

모바일 영상통화 환경에 적합한 하이브리드
변환을 이용한 객체 기반 비디오 코덱 설계



2010년 2월

부 경 대 학 교 대 학 원

컴 퓨 터 공 학 과

전 성 혜

전성혜의 공학석사 학위논문을 인준함

2010년 2월 25일



주 심 공학박사 정 목 동 (인)

위 원 공학박사 이 석 환 (인)

위 원 공학박사 권 기 룡 (인)

목 차

I. 서 론	1
II. 관련 연구	5
2.1 모바일 영상통화 상에서의 비디오 코덱	5
2.2 제로트리 기반의 웨이블릿 코더	8
2.3 기존 모바일 영상통화 서비스의 문제점	13
III. 제안한 모바일 영상통화 비디오 코덱	15
3.1 객체 분리	16
3.2 중요 객체 부호화	19
3.3 비중요 객체 부호화	22
3.4 Inter coding 수행	29
3.5 중요객체에 대한 SNR 스케일러빌리티	33
3.6 데이터 구조 설계	35
IV. 실험 결과 및 분석	43
V. 결론	52
참고문헌	54

표 목 차

표 1. H.264 코덱의 NAL_unit_type 정의	36
표 2. NAL 헤더 내 NAL_ref_idc 재정의	37
표 3. NAL 헤더 내 NAL_unit_type 재정의	37
표 4. 제안한 방법과 H.263 방법에 대한 평균 PSNR 비교	44
표 5. SNR 스케일러빌리티 모드에서 H.263과 중요 객체의 화질 비교	51



그 립 목 차

그림 1. 3GPP TS 26.111 규격 상에서 모바일 영상통화를 위한 비디오 코덱 영역	6
그림 2. 웨이블릿 영역에서 대역 간 계수들의 상관관계	10
그림 3. 기존 서비스되고 있는 모바일 영상통화의 화질 열화 문제점인 (a) 배경 조도에 의한 화질 열화, (b) 빠른 변화 성분에 의한 영상 손실 및 (c) 손떨림 현상에 의한 영상 손실	13
그림 4. 제안하는 비디오 코덱의 부호화 블록도	15
그림 5. 첫 번째 입력 프레임의 (a) 원 영상, (b) 이진마스킹된 영상 및 분리된 (c) 중요 객체와 (d) 비중요 객체	17
그림 6. 중요 객체에 대한 패딩 예; (a) 중요 객체의 경계 매크로 블록 및 (b) 패딩된 중요 객체	17
그림 7. 비중요 객체에 대한 패딩 예; (a) 인접 픽셀 패딩 및 (b) 패딩된 비중요 객체	18
그림 8. 중요 객체의 재 정렬; (a) 재 정렬된 중요 객체에 대한 매크로 블록 및 (b) 중요 객체의 오프셋 값 저장	20
그림 9. Lifting scheme의 구조	23
그림 10. 첫 번째 프레임의 비중요 객체에 대한 4-level DWT 변환 결과	24

그림 11. 3-Level DWT 영역 상에서의 Raster scan order	26
그림 12. 이전 프레임 생성을 위한 복원된 (a) 중요 객체와 (b) 비중요 객체 및 (c) 복원된 이전 프레임	30
그림 13. 첫 번째 프레임과 두 번째 프레임 간의 intercoding 수행 예; (a) 예측 프레임, (b) 현재 프레임, (c) 오차 프레임, (d) 오차 프레임으로부터 분리된 중요 객체 및 (e) 비중요 객체	32
그림 14. SNR 스케일러빌리티가 추가된 중요 객체 부호화 블록도	33
그림 15. 잉여 프레임 생성 예; (a) 원본 중요 객체, (b) 복원된 중요 객체 및 (c) 생성된 잉여 프레임	34
그림 16. H.264의 NAL 구조	36
그림 17. H.264의 NAL 헤더 구조	36
그림 18. IDR_SO/non IDR_SO NAL unit 구성; (a) SO NAL unit 구조, (b) IDR_SO NAL 헤더 및 (c) NonIDR_SO NAL 헤더	38
그림 19. IDR_IO/non IDR_IO NAL unit 구성; (a) IO NAL unit 구조, (b) IDR_IO NAL 헤더, (c) Non IDR_IO NAL 헤더 및 (d) IO Decoding Info.	39
그림 20. DI NAL unit 구성; (a) DI NAL 구조, (b) DI NAL 헤더 및 (c) 데이터 비트 필드	40
그림 21. SOOD NAL unit 구성; (a) SOOD NAL 구조, (b) SOOD NAL 헤더 및 (c) 오프셋 데이터 구성 예	41

그림 22. MVI NAL unit 구성; (a) MVI NAL 구조 및 (b) MVI NAL 헤더 ..	42
그림 23. ELSO NAL unit 구성; (a) ELSO NAL 구조 및 (b) ELSO NAL 헤더 ...	42
그림 24. 0.204bpp에서 복호된 비디오의 프레임별 PSNR 비교	45
그림 25. 0.166bpp에서 복호된 비디오의 프레임별 PSNR 비교	45
그림 26. 0.144bpp에서 복호된 비디오의 프레임별 PSNR 비교	46
그림 27. 0.129bpp에서 복호된 비디오의 프레임별 PSNR 비교	46
그림 28. 0.204 bpp에서 (a) 제안한 방법 및 (b) H.263으로 복호된 비디오의 3번째 프레임	48
그림 29. 0.166 bpp에서 (a) 제안한 방법 및 (b) H.263으로 복호된 비디오의 25번째 프레임	48
그림 30. 0.144 bpp에서 (a) 제안한 방법 및 (b) H.263으로 복호된 비디오의 50번째 프레임	49
그림 31. 0.129 bpp에서 (a) 제안한 방법 및 (b) H.263으로 복호된 비디오의 90번째 프레임	49
그림 32. SNR 스케일러빌리티 모드에서 첫 번째 프레임의 중요 객체 화질 비교를 위한 제안한 방법의 (a) 하위 계층과 (b) 고위 계층 및 (c) H.263 방법으로 복호된 프레임	51

Design of Object-based Video CODEC for the Mobile Video Telephony
Using Hybrid Transform

Sung-Hye Jeon

Department of Computer Engineering, Graduate School,
Pukyong National University

Abstract

Recently, many people have contacted various mobile multimedia services through the mobile terminal owing to the commercialization of 3G communication technology. With the development of processing time on the mobile terminal, the video telephony has been the standout service in mobile multimedia services of 3G mobile communication. Since the service of video telephony has been provided in 3G cellular phone, 3G cellular users can use this service anywhere and anytime. But the quality of the serviced video telephony has been not good yet by the actual mobile restrictions. The user complaint about the low quality of video telephony has been raised incessantly.

There are two major factors of mobile restrictions that limit to the high-quality of the video telephony in the mobile environment. The first one is the streaming bandwidth. Though the streaming bandwidth has been improved somewhat by the commercialization of WCDMA(wideband code division multiple access), HSDPA(high speed downlink packet access) and WiBro(wireless broadband), it is naturally that the bandwidth of the wireless communication is narrow more than that of the wired

communication. The second one is the performance of video CODEC. The video CODEC for the service of video telephony must compute encoding and decoding at the same time. And it must process the massive video stream through software and hardware. The performance of video CODEC should be improved for the high-quality service in the existing mobile video telephony environment.

The video CODEC used in the mobile video telephony environment leads to two quality problems. The first thing is the degradation in the face region under an illumination. The quality in the face region should be guaranteed in video telephony because each other speak over cellular phones seeing themselves. The second thing is the degradation in the background region with the variation. Sometimes the whole region in a frame can be greatly varied by the camera shake. When frames are encoded through video CODEC, the bit rate will be increased in the greatly varied background region and also the serious blocking artifact will be happened in that region.

For solving the above-mentioned quality problems, this paper presents the design of the object-based video CODEC using hybrid transform in mobile video telephony. The proposed design firstly segment each frame into a significant object and an insignificant object and then perform separately the white valance in two objects for alleviating the quality degradation by illumination. Considering the video telephony, a significant object is set to the face region of caller and an insignificant object is set to other region except for the face region, so-called the background region. The proposed design is to improve the quality of a significant object by limiting the bit rate of a insignificant object while minimizing the quality degradation in a greatly varied background region. Thus, we compress a significant object with high quality and low compression ratio and compress an insignificant object with low quality and high compression ratio. Furthermore, we control the bit rate of the video stream in the limited bandwidth by adjusting the compression ratio of each object.

From experimental results, we confirmed that our method has more higher quality than methods in the conventional CODECs at the significant region on the same bit rate.

I. 서론

최근 WCDMA(wideband code division multiple access)를 기반으로 한 HSDPA(high speed downlink packet access) 기술의 상용화로 본격적인 3세대 이동통신 서비스가 제공되고 있다. 3세대 이동 통신 기술은 크게 북미 중심의 동기식 IMT(international mobile telecommunication)-2000 계열과 유럽 중심의 비동기식 IMT-2000 계열로 나뉘어진다[1]. 동기식 IMT-2000 계열은 IS-95 A/B에서 시작하여 3세대 이동통신 기술로서 CDMA 1x, CDMA 1x EV-DO(evolution data only)에서 EV-DO rA로 발전되었다. EV-DO는 기존 CDMA2000 1x 기술에서 고용량의 데이터만을 송수신할 수 있는 기술이나, 단지 데이터만 전송이 가능하므로, 화상 통화나 대용량 서비스가 불가능하다. 비동기식 IMT-2000 계열은 WCDMA에서 현재 상용화가 완료된 HSDPA 기술로 발전되었다. WCDMA는 이론적으로 초당 최대 2Mbps 데이터 전송률을 가지나, 실제로는 300~400kbps로 데이터를 전송 받을 수 있다. HSDPA는 WCDMA를 확장한 고속패킷 통신 규격으로서 이론적으로는 최대 14Mbps의 전송 속도와 실제 2~3Mbps의 속도로 데이터를 전송 받을 수 있다[2]-[4]. WCDMA 보다 7배 정도 향상된 전송 속도를 기반으로 동기식 IMT-2000 계열 기술에서는 제공하기 어려웠던 다양한 서비스가 가능해 졌다. 국내 이동통신사의 경우 SK 텔레콤과 KTF가 비동기식 IMT-2000의 사업자로 선정되어, 모바일 영상통화, 글로벌 로밍, 초고속 데이터 전송이 가능한 데이터 서비스, 교통카드 등의 생활 인프라 서비스 그리고 모바일 뱅킹 등 금융 인프라 서비스를 가능하게 하는 USIM(universal subscriber identity module) 카드 기반 서비스 등 다양한 서

비스를 활발히 제공하고 있다.

이와 같이 HSDPA 기술의 데이터 전송속도의 향상으로 3.5세대 이동통신 서비스 중 가장 두각을 나타내는 서비스가 영상 서비스이다. 영상 서비스는 방송, VOD(video on demand), 멀티미디어 메일 서비스 등의 단방향 영상 서비스와 영상통화와 같은 양방향 영상 서비스로 구분된다. 단방향 영상 서비스는 DMB(digital multimedia broadcasting)나 영상 메일 서비스 등 비디오 데이터의 일방적인 전송이나 수신을 위한 서비스를 말하며, 양방향 서비스는 영상 통화와 같이 비디오 데이터의 수신 및 전송이 동시에 이루어지는 서비스를 말한다. 모바일 환경에서 방대한 데이터 처리량을 갖는 영상 서비스를 제공하기 위하여 반드시 수반되어야 하는 기술이 바로 비디오 코덱 기술이며, 여러 종류의 비디오 코덱이 상용화 되어 있다 [5]-[10]. 단방향 영상 서비스의 경우, DMB 방송 서비스에서는 MPEG-4 AVC(advanced video coding, H.264) 코덱을 사용하고 있으며, VOD 서비스에서는 MPEG-4 Visual과 H.264가 사용된다. 동영상 메시지와 같은 MMS(multimedia message service) 서비스에서도 MPEG-4 Visual을 사용하고 있다. 양방향 영상 서비스를 위한 비디오 코덱으로는 H.261, H.263 코덱이 사용 중에 있다.

모바일 상에서 다양한 영상 서비스가 제공되며 빠르게 발전하고 있지만 실제로 비디오 시스템에 대한 안정성 및 성능은 만족스러운 수준에 이르지 못하고 있다. 원활한 영상 서비스를 제공하는데 제약이 되는 것은 크게 두 가지가 있으며, 첫 번째가 스트리밍 대역폭이고, 두 번째가 비디오 CODEC의 성능이다. 스트리밍 대역폭은 현재 WCDMA, HSDPA, WiBro의 상용화로 인하여 어느 정도 해소되고 있으나, 무선 환경 자체가 유선에 비해 협대역임은 부정할 수 없는 현실이다. 비디오 코덱의 성능은 영상 서비스의 품질을 좌우하는 결정적인 요소로서, 연속된 방대한 비디오 스트림의 처리

가 필요하므로 하드웨어적이든 소프트웨어적이든 막대한 데이터 프로세싱을 요구한다. 하지만 이동 단말기의 배터리 용량의 문제로 인한 비디오 코덱의 저 전력 설계로 복잡도가 높은 고성능의 비디오 코덱을 적용하는 것이 현실적으로 어려운 실정이다. 특히, 영상통화 서비스는 비디오 데이터에 대한 부호화와 복호화가 동시에 이루어져야하며, 실시간 통화를 위해 매우 짧은 지연시간이 요구되는 더욱 제한적인 사항을 갖는다. 배터리 용량, 단말기 탑재 칩 기술의 저 전력화 등의 문제가 지속적으로 연구되고, 발전하고 있지만, 비약적 발전이 이루어지지 않는 한 현재의 문제점들을 해결하기는 어려울 것으로 보인다. 따라서 현재 모바일 환경의 제한적 환경 내에서 향상된 품질의 영상 서비스를 제공하기 위해서는 소프트웨어의 방법적인 개선, 즉 비디오 코덱의 개선이 필요한 실정이다.

현재 상용화된 모바일 영상통화 상에서 가시적으로 나타나는 문제점으로는 배경조도에 의한 화질열화와 프레임 내 큰 변화 성분에 의한 화질열화 발생이다. 본 논문에서는 이와 같은 문제점들을 향상시키기 위하여 객체기반의 하이브리드 변환을 적용한 모바일 비디오 코덱을 설계한다. 제안한 코덱에서는 입력된 모든 비디오 프레임을 중요 객체와 비중요 객체로 분리한 후, 두 객체에 대하여 각각 다른 압축 절차를 수행한다. 모바일 영상통화 환경을 고려하여 중요객체는 얼굴 영역으로 설정되며, 비중요 객체는 얼굴 영역을 제외한 나머지 영역으로 설정된다. 이와 같은 객체 분리 코딩은 중요객체에 대해 독립적인 AWB(auto white balance) 기술을 적용 가능하며, 배경조도에 의한 얼굴 영역의 화질 열화 성분을 향상시킬 수 있다. 또한 객체별 비트율 조정에 의하여 비중요 객체에 해당되는 배경 영역에 많은 비트가 할당되는 문제점이 해결될 수 있다.

제안한 비디오 코덱에서는 얼굴 영역의 중요 객체에 대하여 DCT(discrete cosine transform) 기반의 저 압축을 수행하여 많은 비트를

할당하고, 배경 영역의 비중요 객체에 대하여 DWT(discrete wavelet transform) 기반의 고 압축을 수행하여 작은 비트를 할당함으로써 기존 모바일 네트워크 환경의 제한된 대역폭의 비트율을 만족하도록 한다. 여기서 각 객체의 비트율과 화질은 상호보완(trade-off) 관계이다. 따라서 기존 비디오 코덱의 동일 비트율 기준으로 중요 객체 영역에서는 높은 화질을 가지며, 비중요 객체 영역에서는 낮은 화질을 가진다. 고압축이 수행될 비중요 객체 영역에 기존 DCT 기반 블록 단위 압축방법을 적용할 경우, 블록화 현상이 현저하게 나타날 것이다. 따라서 제안한 비디오 코덱에서는 DWT 기반 프레임 단위 압축 기법을 적용함으로써 기존 영상통화 환경에서 배경영역의 급격한 변화 성분에 의한 블록화 현상을 제거한다.

본 실험에서는 제안한 비디오 코덱의 성능 평가를 위하여 기존 영상통화 환경에서 사용되고 있는 비디오 코덱과 성능 비교를 수행하였다. 성능 평가에서는 동일한 비트율 상에서 제안한 코덱으로 복호된 프레임과 기존 코덱으로 복호된 프레임의 화질을 비교하였다. 실험 결과로부터 전체 프레임에 대한 화질은 기존 코덱에 비하여 낮게 나타났지만 중요 객체 영역의 화질은 기존 코덱보다 우수함을 확인할 수 있었다.

본 논문의 구성을 살펴보면, 2장에서는 현재 영상통화 상에서 사용되는 비디오 코덱과 배경 객체에 적용된 제로트리 기반의 웨이블릿 코더에 대하여 간략히 알아본 다음, 현재 모바일 영상통화 상에서 나타나는 문제점에 대하여 분석한다. 3장에서는 제안한 비디오 코덱의 압축 절차에 대하여 단계별로 살펴본 후, 인코딩과 디코딩을 위해 정의된 데이터 구조에 대해 살펴본다. 4장에서는 실험 결과를 통하여 제안한 방법의 성능을 검토하여 고찰한 뒤, 마지막 5장에서는 본 논문의 결론을 맺는다.

II. 관련 연구

본 장에서는 먼저 제안하는 비디오 코덱의 설계 시 고려사항 및 기존 비디오 코덱의 기술적인 요소를 파악하기 위해 현재 영상통화 서비스에서 사용되고 있는 비디오 코덱에 대하여 살펴본다. 다음으로 비중요 객체의 압축에 적용한 제로트리 기반의 웨이블릿 코더에 대하여 기술한다.

2.1 모바일 영상통화 상에서의 비디오 코덱

모바일 환경은 동일 대역폭에서 유선망에 비해 높은 비용이 소요되며, 모바일 전용 단말기는 저 전력 설계로 높은 성능을 확보하기 어려운 제한적 사항을 가지고 있다. 또한 영상통화와 같은 양방향 서비스는 부호화기와 복호화기의 동시 동작이 요구되며, 지연시간이 매우 짧아야 한다. 이러한 사항을 고려하여 현재 영상통화 서비스는 3GPP(3rd generation partnership project)에서 표준화 한 3G-324M 규격을 따르고 있다[11]~[14]. 3G-324M 규격은 3G 모바일 네트워크에서 영상통화를 위한 프로토콜로서 기존 H.324 규격을 바탕으로 무선 환경에 맞도록 변경, 삭제, 보완된 규격이다. 3G-324M 규격 관련 3GPP TS(technical specification)를 그림 1에 나타내었다. 3G-324M 규격에서는 비디오 데이터의 코딩을 위해 그림 1에서와 같이 H.261(optional)[5]과 H.263(baseline)코덱[7]을 사용하도록 하며, 현재 대부분의 영상통화 상에서 사용되는 비디오 코덱은 H.263을 기반으로 한다.

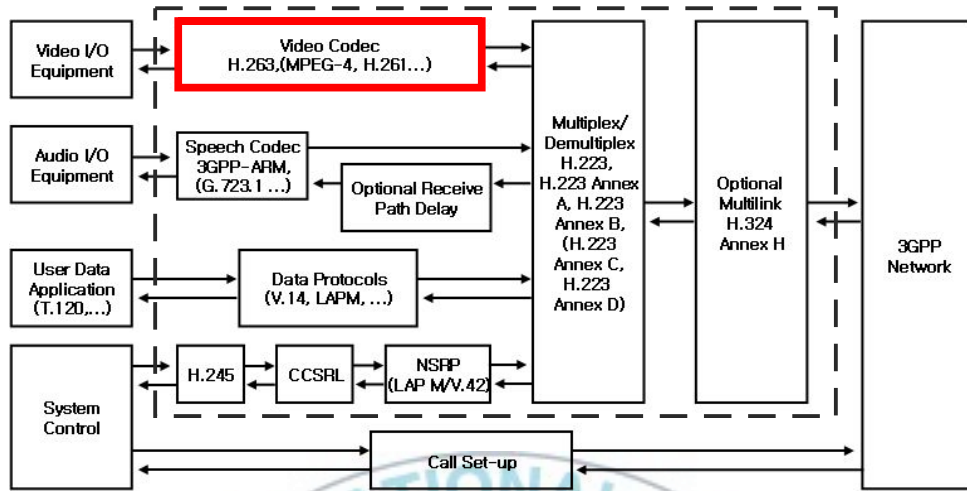


그림 1. 3GPP TS 26.111 규격 상에서 모바일 영상통화를 위한 비디오 코덱 영역

2.1.1 H.261

H.261은 ITU-T(telecommunication standardization)에서 1989년 처음으로 발표된 비디오 압축 표준으로 영상전화 및 영상회의를 목표로 하였다. 전송속도 64kbps의 배수를 지원하는 ISDN(integrated services digital network)을 대상으로 설계되어 Px64로 불리기도 한다. H.261의 특징을 살펴보면 다음과 같다. H.261은 입력 비디오 포맷으로서 CIF(common intermediate format) 또는 QCIF(quarter CIF) 해상도를 가지도록 하며, 양방향성을 고려하여 압축 및 복원과정에서 추가적인 지연이 발생되지 않는 기술들만 적용하여 부호화 지연시간을 최대 150ms 이내로 규정하였다. H.261에서는 움직임 보상 기반의 프레임 간 예측 방법으로서 I 프레임과 P 프레임만을 사용하는 단방향 예측이 수행된다. DCT가 수행되는 블록의 크기는 8×8로 제한하였으며, 움직임 탐색 단위는 16×16 크기의 블록으로 제한하였다. 또한 1화소 정밀도를 갖는 움직임 탐색 기술이 적용되었다.

2.1.2 H.263

H.263 역시 H.261과 마찬가지로 저 전송률을 가지는 통신선로(64kbps 이하)에서 영상전화나 영상회의를 위한 비디오 압축 표준이다. 기본적인 코딩 알고리즘은 H.261과 유사하지만 성능개선을 위해 다음 네 가지의 선택부호화 방법이 첨가되었다.

- Unrestriction motion vector
- Syntax based adaptive arithmetic coding
- Advance prediction
- Bidirectional(forward/backward) frame prediction

H.261에 비하여 같은 화질의 영상에 대해 거의 50% 정도의 압축률을 향상하였다. H.263이 H.261에 비해 개선된 부분을 정리하면 다음과 같다. 우선 각 매크로 블록의 움직임 벡터를 부호화 하는데 있어서 이웃하는 매크로블록의 움직임벡터와 상관도가 높음을 감안해 세 벡터의 중간 값을 취하는 보다 효율적인 방법을 사용하고 있다. 이 방법으로 약 10%의 데이터 감축이 얻어진다. 또한 한 매크로 블록 내에서 움직임을 세분화하는 블록별 움직임 추정이 가능하다. DCT 변환계수의 효율적 양자화를 위해 양자화기를 개선해 약 3%의 데이터절약을 얻고, 또한 양자화 된 변환계수들의 가변길이 부호화는 H.261, JPEG, MPEG1, MPEG2의 2차원 부호를 개선해 화면 내, 화면 간 정보까지를 고려한 3차원부호를 사용한다. 여기서 약 5% 정도의 데이터를 절약할 수 있다. 비트열의 선택스(syntax)에 있어서도 기존의 H.261보다 크게 단순화해 순수한 정보 비트이외의 오버 헤드를 줄이고 있다. 이 밖에도 선택스 기반 적응 산술부호화와 양방향 프레임 예측기

법 등 성능향상을 가져오는 기술이 새롭게 도입되었으나, 복잡도가 높아 사용여부를 옵션으로 설정 할 수 있도록 하였다. 선택스 기반 적응 산술부호화를 사용할 경우 약 5~14%정도의 효율을 증가 시킨다.

2.2 제로트리 기반의 웨이블릿 코더

기존 블록 단위의 DCT 변환 기반의 부호화 방식은 블록의 경계에서 발생하는 블록화 현상(blocking effect)으로 인하여 화질 열화 성분이 발생된다. 이와 달리 웨이블릿 변환은 공간영역과 주파수영역에서 국부화된 기저(basis)를 이용하여 인간의 시각적 특성에 맞게 분해하고 부호화함으로써 주관적 화질의 향상을 가져온다. 웨이블릿 기반의 코더에 대한 초기 연구는 Antonini 등[15]이 시작하였으며, 주로 웨이블릿 계수의 통계적인 성질의 차이점을 강조하는 연구를 진행하였다. 이후, Lewis와 Knowles[16]가 낮은 주파수와 높은 주파수의 계수들의 제로트리를 예측하는 알고리즘을 발표함으로써 웨이블릿 기반 코딩 연구의 새로운 전기를 마련하였다. 제로트리를 이용한 압축 방법은 웨이블릿 변환된 영상의 계수 값이 동일한 방향을 갖는 대역 사이에서 상관관계를 갖는다는 점을 이용한다. 널리 사용되고 있는 제로트리 기법으로 Shapiro[17]가 제안한 EZW(embedded zerotree wavelet)와 Pearlman 등[18]이 제안한 SPIHT(set partitioning in hierarchical trees)를 들 수 있다. EZW는 제로트리를 이용하여 웨이블릿 계수를 점진적으로 전송하는 방법이며, SPIHT는 기본적인 접근 방법은 EZW와 동일하나 제로트리의 부호화 부분을 개선하였다. EZW와 SPIHT는 다음의 세 가지 기본 개념을 기반으로 한다.

- ① 제로트리 구조를 이용하나 0이 아닌 계수 위치를 효율적으로 부호화 한다.
- ② 계수를 절대값의 크기에 따라 정렬한다.
- ③ 계수의 크기를 나타내는 비트는 MSB(most significant bit)이다.

EZW와 SPIHT의 특징은 각 웨이블릿 계수들을 균일 양자화한 후 비트 플레인(bit-plane)을 이용하여 전송하는 것이다. 즉 주어진 순서에 따라 각 계수들을 탐색하고, 중요하다고 판명된 계수의 비트를 MSB 순으로 전송한다. 웨이블릿 코더의 대표적인 방법인 EZW와 SPIHT에 대하여 아래에 자세히 살펴보도록 한다[19],[20].

2.2.1 EZW 기법

EZW 알고리즘에서는 웨이블릿 변환된 영상의 계수들을 해당 대역과 방향성에 따라 중요도 지도(significance map)를 구성하고, 대역 간 계수들의 중요성 여부에 따라 제로트리 부호화를 수행한다. 대역 간 웨이블릿 계수들의 상관관계는 다음 그림 2와 같이 쿼드 트리(quad-tree) 구조를 갖는다.

제로트리를 이루는 계수들의 관계를 표현하기 위해 다음과 같이 용어를 정의한다. 다해상도(multiresolution) 분해된 웨이블릿의 상위 계층에 위치한 계수를 부모(parent)라 하며, 하위 계층에서 동일한 위치와 방향에 있는 계수를 자손(children)이라 한다. 또한 동일한 위치와 방향에 있는 모든 하위 계층의 계수들을 후손(descendants)이라 정의한다. 최상위 계층의 최저주파수 대역과 최하위 계층을 제외한 모든 계수들을 네 개의 자손 계수를 갖는다. 대역간의 상관관계를 이용한다는 것은 ‘만약 한 계수가 주어진 문턱치보다 작으면, 그 계수의 후손들 역시 작은 값을 가질 것이다’라고 가정하는 것을 의미한다.

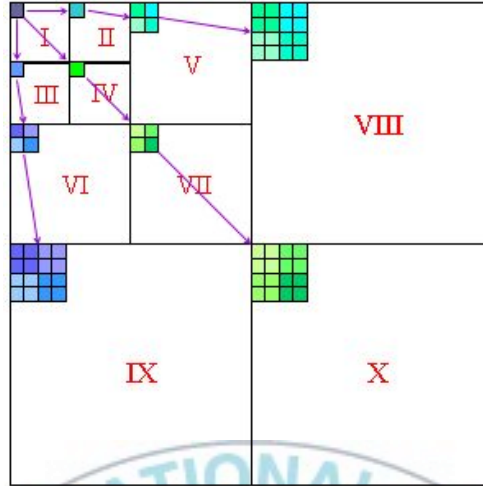


그림 2. 웨이블릿 영역에서 대역 간 계수들의 상관관계

제로트리에는 4가지의 기호가 사용된다. 만약 주어진 문턱치 T_i 에 대하여 웨이블릿 계수 x_i 의 절대값이 T_i 보다 크다면 이 계수는 중요하다고 판명되며, POS(positive) 또는 NEG(negative)로 부호화된다. 여기서 x_i 가 중요하다고 판명된 계수의 자손이면서 x_i 의 모든 후손이 중요하지 않다고 판명된다면, 이 x_i 는 ZTR(zerotree root)라고 한다. 웨이블릿 변환의 특성에 의해 중요하지 않은 계수의 모든 후손은 중요하지 않을 확률이 매우 높으므로, 많은 수의 웨이블릿 계수들을 하나의 ZTR 심벌로 부호화 할 수 있다.

EZW 알고리즘은 연속근사 양자화(successive approximation quantization) 방법을 이용한다. 연속근사 양자화 방법에서 계수의 중요도를 판명하기 위해 일련의 문턱치 $T_0, T_1 \cdots T_{n-1}$ 을 적용시킨다. 이러한 문턱치들은 $T_i = T_{i-1}/2$ 의 관계를 가진다. 초기 문턱치 T_0 는 모든 계수 x_i 에 대하여 $|x_i| < 2T_0$ 이다.

2.2.2 SPIHT 기법

SPIHT는 앞서 설명한 EZW와 동일하게 제로트리를 이용한 웨이블릿 계수의 점진적 전송 방법이다. EZW와의 주요 차이점은 웨이블릿 계수를 계수의 중요도에 따라 부분 집합으로 분할해가는 과정과 분할된 계수들을 부호화하는 과정이다. 전체 중요도 지도는 모든 문턱치에서 다음의 세 가지 리스트를 통하여 구해진다.

- LSP(list of significant pixels)
- LIP(list of insignificant pixels)
- LIS(list of insignificant sets)

전체 알고리즘은 다음 네 가지 단계로 이루어진다.

- 초기화
- Sorting Pass
- Refinement Pass
- 양자화 값 갱신

다음과 같이 정의된 웨이블릿 계수들의 집합이 제로트리 부호화를 위해 사용된다.

- $O(i,j)$: 계수 (i,j) 에 대한 모든 offspring 계수의 좌표로 이루어진 집합
- $D(i,j)$: 노드 (i,j) 에 대한 모든 descendants 노드의 좌표로 이루어진 집합
- $L(i,j) = D(i,j) - O(i,j)$

- H : 가장 높은 피라미드 레벨, 즉 가장 낮은 해상도를 갖는 밴드에 있는 모든 계수들의 집합

여기에서 offspring은 (i,j) 와 동일한 방향과 위치에 있는 하위 계층에 있는 모든 계수를 말하며, descendants는 offspring에서 (i,j) 의 바로 밑의 계층에 있는 4개의 계수를 제외한 모든 계수의 집합을 뜻한다.

다음과 같은 세 가지의 규칙이 집합 분할(set partitioning) 방식에 적용된다.

- 초기 분할은 $(i,j) \in H$ 인 모든 (i,j) 와 $D(i,j)$ 로 이루어진다.
- 만약 $D(i,j)$ 이 중요하다면, 즉 주어진 문턱치보다 큰 화소가 존재한다면, $D(i,j)$ 는 $L(i,j)$ 와 네 개의 단일 화소 집합들인 $(k,l) \in O(i,j)$ 로 나누어진다.
- 만약 $L(i,j)$ 가 중요하다면, $L(i,j)$ 는 네 개의 집합 $D(k,l)$ 로 나누어진다. 여기서 $(k,l) \in O(i,j)$ 이다.

웨이블릿 계수들의 좌표로 이루어진 집합 Ω 의 중요도 여부를 나타내는 함수 $S_n(\Omega)$ 는 다음과 같이 정의된다.

$$S_n(\Omega) = \begin{cases} 1, & \text{if } \max_{(i,j) \in \Omega} \{|C_{i,j}|\} \geq 2^n \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

2.3 기존 모바일 영상통화 서비스의 문제점

3세대 이동통신 서비스로 대표되는 영상통화 서비스에 대해 향상된 품질을 제공하기 위해 기존 모바일 영상통화 상에서 나타나는 대표적인 문제점을 분석하였다. 그림 3은 3G 휴대폰(IM-S240K)으로 영상통화 후 저장된 동영상을 캡처(capture)한 화면이다.

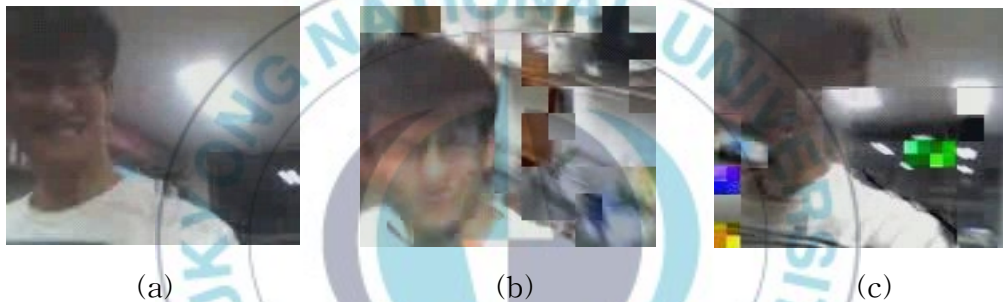


그림 3. 기존 서비스되고 있는 모바일 영상통화의 화질 열화 문제점인 (a) 배경 조도에 의한 화질 열화, (b) 빠른 변화 성분에 의한 영상 손실 및 (c) 손떨림 현상에 의한 영상 손실

그림 3의 (a)에는 배경 조도에 의해 화질 열화 성분이 발생하는 것을 볼 수 있다. 영상통화 환경에서는 서로의 얼굴을 보고 대화하기 때문에 얼굴에 대한 화질이 보장되어야 한다. 하지만 그림 3의 (a)에서와 같이 배경 조도에 의해 얼굴의 형태를 알아보기 힘들만큼 화질 열화가 발생하는 것을 확인할 수 있다. 이와 같은 문제를 개선하기 위하여 AWB 기술을 프레임에 적용할 경우, 화면의 전체적인 밝기가 증가하여 프레임 자체가 왜곡될 수 있다. 그림 3의 (b)에서는 빠른 변화 성분에 의해 블록 에러와 같은 영

상 손실이 발생하는 것을 확인 할 수 있다. 모바일 영상통화 사용자들은 단말기를 손으로 들고 통화하는 것이 일반적이다. 이럴 경우, 단말기와 가까운 거리에 있는 얼굴 영역이 프레임 내 전경(foreground)에 위치하게 되며, 상대적으로 거리가 먼 배경(background) 영역은 사소한 움직임에도 그 변화가 크게 나타나게 된다. 기존의 대다수의 DCT 기반 코덱은 프레임간의 상관성을 고려한 시간적 중복정보에 대한 압축을 수행한다. 즉, 이전에 코딩된 프레임과 현재 코딩될 프레임간의 중복되는 데이터 영역을 추정하여, 중복된 영역은 코딩을 수행하지 않고 변화된 영역에 대해서만 코딩을 수행한다. 따라서 기존 비디오 코덱은 제한된 대역폭 내에서 얼굴 영역에 비해 변화량이 큰 배경 영역에 대해 상대적으로 많은 비트를 할당하여 코딩을 수행한다. 또한, 배경 영역에 대해 갑작스런 움직임 등의 이유로 큰 변화량을 가질 때, 중복정보를 추정하기가 어렵게 되며, 그림 3의 (b)에서 볼 수 있듯이 블록 에러와 같은 화질 열화 성분들이 나타나게 된다. 영상통화 환경의 특성 상, 손 떨림 현상 등 프레임 전체에 대해 큰 변화 성분을 가질 수 있으며, 이러한 경우에는 그림 3의 (c)와 같이 프레임 전체에 걸쳐 블록 에러가 발생 할 수 있다. 따라서 본 논문에서는 모바일 영상통화 시 나타나는 이러한 문제점들을 향상시키기 위한 방법을 제안하며, 구체적인 절차와 방법은 3장에서 서술하도록 한다.

Ⅲ. 제안한 모바일 영상통화 비디오 코덱

제안하는 비디오 코덱은 매 프레임마다 중요 객체와 비중요 객체를 분리하며, 중요 객체에 대해서는 저압축, 고화질을 제공하도록 부호화를 수행한다. 이에 반해, 비중요 객체에 대해서는 고압축, 저 화질을 제공하도록 부호화를 수행함으로써 기존의 제한된 대역폭 내에서 비트율을 만족하도록 조절하게 된다. 또한, 부호화 과정에서 각 객체에 적합한 변환 방법을 적용하여 압축 수행 절차를 달리 한다. 중요 객체는 압축효율을 고려하여 매크로 블록 단위의 DCT 기반의 압축 절차를 수행한다. 기존 블록 기반의 DCT를 이용한 비디오 방법은 고압축 수행 시, 심각한 블록 에러가 발생하는 문제점이 나타났다. 따라서 고압축이 수행될 비중요 객체에 대해서는 프레임 기반의 DWT를 이용한 부호화 방법을 적용하였다. 그림 4에 제안하는 비디오 코덱의 부호화 과정을 도식하였다.

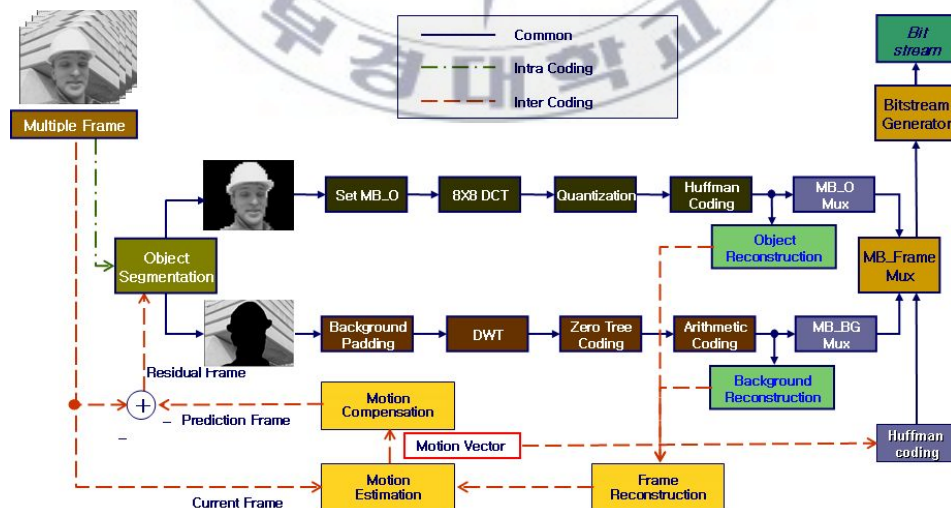


그림 4. 제안하는 비디오 코덱의 부호화 블록도

3.1 객체 분리

본 절에서는 프레임으로부터 각 객체를 분리하는 과정에 대하여 살펴보도록 한다. 모바일 영상통화 시, 얼굴 마주 보며 통화를 하는 것이 일반적이다. 따라서 제안하는 비디오 코덱에서는 영상통화 환경을 고려하여 중요 객체를 얼굴이 포함된 사람 영역으로 설정하였고, 프레임 내에 중요 객체를 제외한 배경 영역을 비중요 객체로 설정하였다. 본 논문에서는 효율적인 압축 방법 및 절차에 초점을 두고 있기 때문에 객체의 분리는 실험의 편의 상 MPEG-4 코어 프로파일(core profile)에서 임의의 형상 코딩(arbitrary shape coding)을 위해 사용된 이진 알파 마스크(binary alpha mask)를 이용하여 수행하였다. 실제로 현재 여러 모바일 폰에서 이미 기술이 상용화된 얼굴 인식 알고리즘의 적용 가능하며 향후 적용할 계획에 있다. 객체 분리 시 매 프레임마다 해당 이진마스크를 가지고 있으며, 그림 5에 이진마스크와 분리된 객체에 대하여 나타내었다. 그림 5의 (b)에 첫 번째 프레임에 대한 이진마스크와 그림 5의 (c)와 (d)에 분리된 각 객체를 나타내었다.

분리된 각 객체는 경계영역에서 큰 고주파 성분을 포함하고 있다. 주파수 영역 상에서 압축이 수행되기 때문에 경계영역에서 심각한 화질열화가 발생할 것이다. 따라서 경계 영역의 고주파 성분에 대한 처리가 필요하다. 중요 객체는 매크로 블록 단위로 DCT를 기반으로 압축이 수행되기 때문에 경계영역에 해당하는 매크로 블록에 대해서 패딩(padding)을 수행한다. 그림 6의 (a)와 같이 중요 객체의 경계 매크로 블록은 중요 객체에 해당하는 불투명 영역과 중요 객체에 해당하지 않는 투명 영역으로 구분 할 수 있으며, 두 영역의 경계에 큰 고주파 성분이 포함되어 있다. 이를 해결하기

위하여 입력 프레임으로부터 투명 영역에 패딩을 수행한다. 그림 6의 (b)에 경계 매크로 블록에 대해 패딩을 수행한 결과를 나타내었다.



그림 5. 첫 번째 입력 프레임의 (a) 원 영상, (b) 이진마스크된 영상 및 분리된 (c) 중요 객체와 (d) 비중요 객체

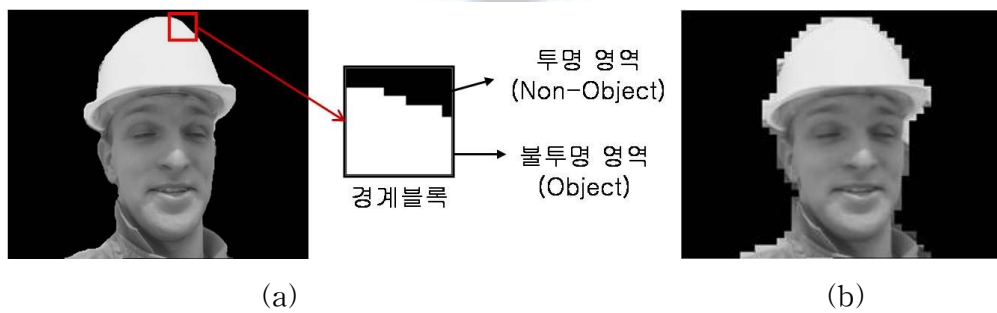


그림 6. 중요 객체에 대한 패딩 예; (a) 중요 객체의 경계 매크로 블록 및 (b) 패딩된 중요 객체

비중요 객체는 프레임 단위로 DWT를 수행한다. 분리된 비중요 객체의 투명영역, 즉 중요 객체가 있던 영역은 인접 픽셀로서 그림 7의 (a)와 같이 패딩이 수행된다. 투명영역의 가운데를 기준으로 인접한 픽셀을 좌측과 우측으로부터 채워 나간다. 그림 7의 (b)에 첫 번째 프레임에 대해 패딩이 수행된 비중요 객체를 나타내었다.

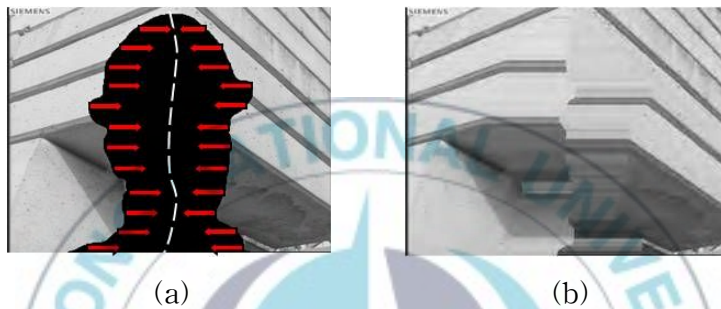


그림 7. 비중요 객체에 대한 패딩 예; (a) 인접 픽셀 패딩 및 (b) 패딩된 비중요 객체

위와 같이 패딩이 수행된 객체는 각각 다른 변환 방법을 적용한 부호화 과정을 거치게 된다.

3.2 중요 객체 부호화

시각적으로 중요하다고 간주되는 얼굴 영역의 객체에 대해서는 객체에 대한 정확한 움직임 추정 및 보상과 고화질의 양자화 파라미터를 적용하기 위해 기존의 잘 알려진 8x8 블록 단위의 DCT 변환을 이용하여 인코딩한다. 중요 객체의 인코딩 절차는 매크로 블록 재 정렬 및 중요 객체의 오프셋 값을 분리하는 매크로 블록 초기화 단계와 DCT 및 양자화 수행 단계, 그리고 허프만 부호화 단계까지 크게 4단계 구성된다. 각 단계별 중요 객체의 코딩 방법에 대하여 살펴보도록 한다.

3.2.1 매크로 블록 초기화

분리된 중요 객체는 효과적인 DCT 수행을 위해 2차원 버퍼에 재 정렬하게 된다. 재 정렬은 좌측 상단에서 우측 하단 방향으로 매크로 블록 단위로 버퍼에 저장되며, 15개의 매크로 블록이 버퍼의 수평 방향으로 위치하도록 정의하였다. 그림 8의 (a)에 실제 버퍼에 재 정렬된 매크로 블록을 나타내었다. 현 단계에서는 재 정렬 외에도 중요 객체의 복호를 위한 매크로 블록의 오프셋 값을 저장하게 된다. 본 코덱에서는 프레임이 내의 매크로 블록의 수평 라인을 하나의 슬라이스(slice)로 정의하였다. 그림 8의 (b)에 매크로 블록 단위로 분할된 프레임과 저장될 오프셋 포인트를 예를 들어 나타내었다. 사각형 하나는 매크로 블록에 해당하며, 음영이 있는 사각형은 중요 객체를 의미한다. 슬라이스 내의 첫 번째 매크로 블록의 시작 포인트와 매크로 블록의 개수를 저장 후 중요 객체의 복호를 위해 전송하게 된다.

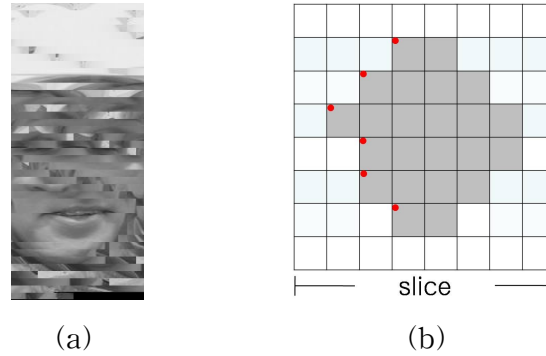


그림 8. 중요 객체의 재 정렬; (a) 재 정렬된 중요 객체에 대한 매크로 블록 및 (b) 중요 객체의 오프셋 값 저장

3.2.2 DCT 변환, 양자화, VLC(variable length coding)

중요 객체는 8x8 매크로 블록 단위로 DCT 기반의 압축절차가 된다. 먼저, 재 정렬된 각 매크로 블록은 DCT를 수행하여 공간영역에서 주파수 영역으로 변환된다. 이후 각 계수 값 X 들은 사람의 눈에 영향이 적은 고주파 성분을 효과적으로 제거하기 위해 양자화가 수행된다. 양자화는 식 (2)을 이용하여 수행되며, 양자화 파라미터 QP 로서 중요 객체의 화질을 결정하게 된다. 양자화를 수행하면 크지 않은 DCT 계수를 제거하여, '0' 값을 많이 포함하는 양자화된 계수들의 배열을 생성한다.

$$FQ = IntegerRound\left(\frac{X}{QP}\right) \quad (2)$$

양자화된 계수 FQ 는 대부분 '0'으로 구성되어 있으므로, 보다 효과적인

엔트로피 코딩을 하기 위해 데이터를 정렬한다. 데이터 정렬은 보편적으로 계산 효율이 높은 DC 계수로부터 지그재그로 정렬하는 지그재그(zigzag) 스캔을 사용한다. 지그재그 스캔은 DCT 코딩된 계수들을 통계적인 기법으로 압축하기 위해 저주파 성분부터 고주파 성분으로 일정한 순서로 배열하기 위한 방법이다. 지그재그 스캔을 통하면 '0'의 계수들이 연속적으로 나열될 수 있으므로 압축 효율을 극대화 할 수 있다. 다음으로 무손실 부호화 방법인 RLC(run length coding)이 수행된다. RLC는 동일한 수가 얼마나 반복되는지를 나타내어 압축하며, 반복된 연속된 수가 많을수록 압축률이 높아지게 된다. 따라서 RLC를 사용하면 연속된 '0'의 계수들을 간결하게 표현 할 수 있다. RLC이 수행된 후 엔트로피 코딩으로서 허프만 코딩이 수행된다. 허프만 코딩은 각 심벌에 대한 확률을 구하여 가장 큰 확률을 가지는 심벌에 대하여 짧은 코드 길이를 부여하는 가변 길이 코딩기법이다. 실제 코드에서는 MPEG 표준 권고안에서 정의한 VLC table을 사용하여 RLC와 허프만 코딩을 구현하였다.

3.3 비중요 객체 부호화

기존의 비디오 압축 기법들은 공간적 중복 정보(spatial redundancy), 시간적 중복 정보(temporal redundancy), 통계적 중복 정보(statistical redundancy)를 제거한다. 이 중에서 시간적 중복 정보 제거는 프레임간의 높은 상관관계를 이용하여 중복된 데이터를 제거함으로써 압축률을 높이는 방법을 사용한다. 영상통화 시, 비중요 객체에 해당하는 배경영역은 중요객체인 얼굴 영역보다 뒤에 위치하며, 그 변화가 크게 나타난다. 따라서 중요객체 보다 변화량이 많은 배경영역이 낮은 상관관계를 가지며, 이는 중요하지 않은 영상 정보에 상대적으로 많은 비트수를 할당하는 문제를 가지고 있다. 제안하는 비디오 코덱에서는 상대적으로 비중요 객체로 간주되는 배경영역에 대하여 저화질을 갖도록 고압축을 수행한다. 기존 대부분의 비디오 코딩 기법들은 블록기반의 DCT 변환을 이용하여 압축을 수행한다. 따라서 기존의 방법으로 비중요 객체 대하여 고압축을 수행할 때 심각한 블록 에러가 발생한다. 이를 해결하기 위해 프레임 단위로 DWT 영역 상에서 제로트리 인코딩을 수행한다. 프레임 단위로 인코딩을 수행함으로써 블록 에러와 같은 화질열화를 향상 시킬 수 있다.

3.3.1 웨이블릿 변환

분리된 비중요 객체는 DWT을 통해 주파수 영역으로 변환된다. 일반적으로 프레임 기반의 변환 방법은 블록 기반 변환 방법에 비해 요구되는 메모

리가 크다. 따라서 빠른 수행속도와 모바일 환경을 고려한 정수 연산 및 메모리 사용의 최소화를 위해 식 (3)의 Le Gall 5/3-tab 필터뱅크를 적용하여 그림 9와 같이 리프팅 구조(lifting scheme)를 구현하였다.

$$\begin{bmatrix} S(z) \\ D(z) \end{bmatrix} = \begin{bmatrix} -\frac{1}{8}z^{-1} + \frac{3}{4} - \frac{1}{8}z & \frac{1}{4} + \frac{1}{4}z \\ -\frac{1}{2}z^{-1} - \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} X_e(z) \\ X_o(z) \end{bmatrix} \quad (3)$$

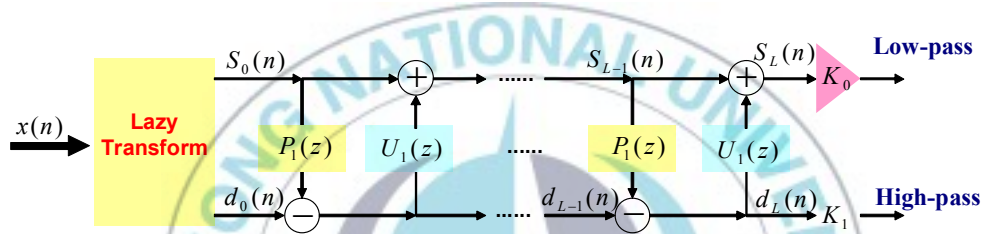


그림 9. Lifting scheme의 구조

첫 번째 프레임의 분리된 비중요 객체에 대한 웨이블릿 변환된 결과를 그림 10에 나타내었다. 웨이블릿 변환된 결과는 다해상도 분석이 가능하며, 부모 자손 관계의 자기 상관성(self-similarity)을 갖는다. 그림 10에서 볼 수 있듯이 영상의 거의 모든 정보가 저주파수 대역이 몰려 있으며, 그 외의 대역에서는 거의 '0'에 가까운 값을 갖고 있음을 알 수 있다. 따라서 비트 플레인 방법을 사용해서 부호화 할 경우에 저주파수 대역 외의 다른 대역들은 거의 '0'인 관계로 비트 플레인 전체를 부호화 하는 것은 상당히 비효율적이 되며, 따라서 각각의 비트 플레인을 부호화 할 때 모든 비트 플레인을 부호화 하지 않고, ZTR 방법을 사용해서 효율적으로 부호화 하게 된다.

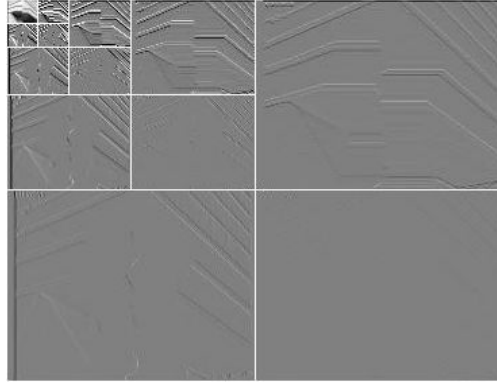


그림 10. 첫 번째 프레임의 비중요 객체에 대한 4-level 웨이블릿 변환 결과

3.3.2 제로 트리 코딩

앞서 말한 바와 같이 웨이블릿 변환된 계수 값들은 저주파 서브 밴드의 '뿌리(root)'에서 시작하는 '트리(tree)' 형태로 간주할 수 있다. 즉, LL 밴드에 존재하는 단일 계수는 하위 레이어의 네 개의 계수와 대응하는 쿼드트리 구조의 부모-자손 관계를 갖는다. 또한, 하위 레이어에 해당하는 고주파 서브 밴드에 있는 많은 계수들이 대부분 작은 값을 가지고 있기 때문에 영상 화질의 큰 손실 없이 양자화 될 수 있다. 제로 트리 코딩은 이러한 특성을 이용하였으며, 비트율-양자화 왜곡 성능이 좋고 계수 값이 큰 중요 계수부터 먼저 부호화시켜 전송하는 임베딩(embedding) 특성을 가지고 있기 때문에 비트율 제어가 가능하다. 제안하는 비디오 코덱에서는 구성파일 내에 비중요 객체에 대한 비트율을 설정 가능하도록 하였으며, 식 (4)에 의해 비트율은 바이트 크기로 변환되어 정해진 크기만큼만 코딩하도록 한다.

$$ByteBudget = \frac{H_{fm} \times V_{fm} \times T_{bpp}}{8} \quad (4)$$

여기서 H_{fm} 은 입력 프레임의 수평 해상도를 말하며, V_{fm} 는 입력 프레임의 수직 해상도를 말한다. T_{bpp} 는 미리 설정한 원하는 비트율을 의미한다.

적용한 제로 트리 코딩의 자세한 부호화 과정은 다음과 같다. 부호화 과정은 크게 주부호화과정(dominant pass)과 종속부호화과정(subordinate pass)으로 나누어지게 되는데, 양자화 계수를 반으로 줄여나가면서 이 두 과정을 반복하게 된다. 주부호화과정에서는 양자화 계수와 웨이블릿 계수를 비교해서 웨이블릿 계수의 중요도를 심별로 정의한다. 이때 중요도 판별 및 심별은 다음과 같이 결정한다. 부모계수의 값이 클 경우 자손계수의 값 또한 클 확률이 높으며, 부모계수의 값이 작을 경우 자손계수의 값 또한 작을 확률이 높다. 따라서 적용된 제로트리 코딩은 임계값을 설정하여 현재 부호화 하려는 계수의 절대값이 임계값 보다 큰지 작은지를 판단하고, 자손계수들 중에 절대값이 임계값 보다 큰 계수가 존재하는지를 판단한다. 다음으로 현재 부호화 하려는 계수가 임계값 보다 클 경우에 항상 자손계수를 부호화하고, 그렇지 않을 경우에는 자손계수의 상황에 따라서 자손계수를 부호화 할 것인지를 결정한다. 중요도 판별 후 계수들은 다음 4개의 심별을 사용해서 부호화 되며, 각 심별의 정의는 다음과 같다.

- POS - Positive Significant, 부호화 하려는 계수의 절대값이 임계값 보다 크며, 계수의 부호가 양수인 경우.
- NEG - Negative Significant, 부호화 하려는 계수의 절대값이 임계값 보다 크며, 계수의 부호가 음수인 경우.

- IZ - Isolated Zero, 부호화 하려는 계수의 절대값이 임계값 보다 작지만, 자손계수들 중에는 임계값 보다 큰 값이 존재할 경우.
- ZTR - Zero Tree Root, 부호화 하려는 계수의 절대값이 임계값 보다 작으며, 모든 자손계수들 역시 임계값 보다 작을 경우.

따라서 현재 부호화 하려는 계수가 POS, NEG와 같은 중요계수로 판단된 경우와, IZ로 판단될 경우에는 항상 자손계수 각각에 대해서 POS, NEG, IZ, ZTR 심별을 사용해서 다시 중요도를 판별하여 부호화하고, ZTR로 판단될 경우에는 더 이상 자손계수를 부호화하지 않는다. 즉 영상의 같은 위치를 복원하는데 관계된 부모, 자손계수 모두를 하나의 블록으로 생각할 때에, ZTR 심별은 현재 부호화하는 계수가 포함된 블록의 자손계수는 모두 0이라는 EOB(end of block)에 해당하는 심별이다.

0	1	4	67	5	6	68	69	9	10	13	14	72	73	76	77
2	3	130	193	7	8	70	71	11	12	15	16	74	75	78	79
25	88	46	109	131	132	194	195	17	18	21	22	80	81	84	85
151	214	172	235	133	134	196	197	19	20	23	24	82	83	86	87
26	27	89	90	47	48	110	111	135	136	139	140	198	199	202	203
28	29	91	92	49	50	112	113	137	138	141	142	200	201	204	205
152	153	215	216	173	174	236	237	143	144	147	148	206	207	210	211
154	155	217	218	175	176	238	239	145	146	149	150	208	209	212	213
30	31	34	35	93	94	97	98	51	52	55	56	114	115	118	119
32	33	36	37	95	96	99	100	53	54	57	58	116	117	120	121
38	39	42	43	101	102	105	106	59	60	63	64	122	123	126	127
40	41	44	45	103	104	107	108	61	62	65	66	124	125	128	129
156	157	160	161	219	220	223	224	177	178	181	182	240	241	244	245
158	159	162	163	221	222	225	226	179	180	183	184	242	243	246	247
164	165	168	169	227	228	231	232	185	186	189	190	248	249	252	253
166	167	170	171	229	230	233	234	187	188	191	192	250	251	254	255

그림 11. 3-Level DWT 영역 상에서의 Raster scan order

중속부호화과정에서는 주부호화에서 중요 계수로 판단된 계수의 양자화의 정밀도를 높여 나간다. 이 때 주부호화과정에서 부호화 하는 순서는 아래 그림 11과 같이 가장 분해가 많이 이루어진 저주파 대역부터 고주파 대역으로 올라가는 순서로 부호화가 이루어진다. 중속부호화과정에서 부호화 하는 순서는 가장 먼저 중요계수로 판단된 계수부터 부호화가 이루어진다.

적용된 제로 트리 코딩의 부호화 과정을 요약하면 다음과 같다.

- ① 모든 계수를 조사해서 최대값을 찾아내어 이 값을 기준으로 양자화 계수를 설정한다.
- ② 주부호화과정에서 부호화가 이루어질 대상을 설정한다. 처음 시작일 경우 모든 계수가 부호화 대상이 된다.
- ③ 주부호화과정을 시작한다. 부호화 방법으로는 제로트리 Root 기법을 사용한다.
- ④ 중속부호화과정을 시작한다. 주부호화과정에서 판단된 중요계수들의 정밀도를 높여나간다.
- ⑤ 임계값을 반으로 낮춘다.
- ⑥ 주부호화대상에서 중요계수로 판단된 모든 계수를 제외시킨다.
- ⑦ ③번의 주부호화과정부터 반복 시행한다.

이렇게 주부호화과정과 중속부호화과정을 통해서 중요 계수를 계속 발생시켜 나가며 양자화의 정밀도를 높여나감에 따라서 부호화나 복호화 시 임의의 위치 어느 곳에서나 멈출 수 있는 특성을 가지고 있어서 비트율을 쉽게 조절할 수 있다.

3.3.3 산술 부호화

엔트로피 코더로서 산술 부호화(arithmetic coding)가 사용되는데, 이것은 주부호화과정에서 심벌이 하나 발생할 때마다 바로 부호화 해나갈 수 있으며, 마찬가지로 종속부호화과정에서도 부호가 하나 발생할 때마다 바로 부호화 할 수 있기 때문에, 연속적인 근사 양자화의 특성을 그대로 살려줄 수 있는 장점이 있으면서도 허프만 부호화와 동등한 성능을 발휘하게 된다. 특히 산술 부호화는 각각의 심벌에 비트를 할당하는 방법이 아니라, 주어진 구간의 범위를 각각의 심벌에 해당하는 범위만큼 좁혀나가면서 부호화 하는 방법으로 적용된 제로 트리 코딩 방법과 같이 부호화 하려는 심벌의 수가 적을 때에 아주 효과적으로 부호화 할 수 있다.

주부호화과정에서는 심벌 4개(POS, NEG, IZ, ZTR)로 구성된 히스토그램(histogram)을 사용해서 부호화하며, 종속부호화과정에서는 심벌 2개(1, 0)로 구성된 히스토그램을 사용해서 부호화 한다. 실제로 주부호화과정 시 부모 계수가 중요계수인지와, 바로 전에 부호화된 계수가 중요계수인지의 서로 다른 4가지 상황에 따라 다른 히스토그램을 사용하게 된다. 따라서 적용된 산술 부호화에서는 심벌 4개로 구성된 4개의 히스토그램과, 심벌 2개로 구성된 한 개의 히스토그램이 사용된다.

3.4 Intercoding 수행

제안하는 비디오 코덱은 압축률 향상을 위해 시간적 중복 정보(temporal redundancy)에 대한 압축을 수행한다. 연속된 프레임은 매우 높은 유사성을 가지므로 이들의 공통점을 찾아내어, 유사한 블록에 대하여 압축을 수행하지 않고 방향 정보 즉, 움직임 벡터(motion vector)만을 전송하는 방법이다. 시간적 중복 정보에 대한 압축을 수행하기 위하여 움직임 추정(motion estimation)과 움직임 보상(motion compensation)이 행해지며, 기존의 DCT 기반 압축방법에서 사용된 움직임 추정 및 보상 방법을 제안하는 비디오 코덱에 적용하였다. GOP(group of picture)를 설정하여, GOP 단위로 intercoding이 수행되도록 하였다. GOP는 구성파일 내에 파라미터를 두어 설정 가능하도록 하였다.

3.4.1 움직임 추정

현재 프레임과 이전의 프레임간의 가장 유사도가 높은 블록을 구함으로서 움직임 추정이 수행된다. 먼저, 이전 프레임을 복원하기 위하여 각 객체가 복원되며, 복원된 각 객체로부터 프레임을 복원 시킨다. 그림 12에 복원된 각 객체와 이전프레임을 나타내었다.

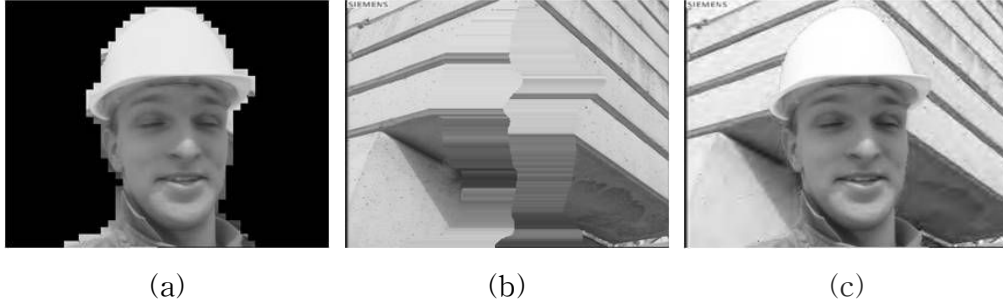


그림 12. 이전 프레임 생성을 위한 복원된 (a) 중요 객체와 (b) 비중요 객체 및 (c) 복원된 이전 프레임

복원된 이전 프레임과 현재 입력 프레임간의 움직임의 변화를 비교하기 위해 16x16 블록별로 매칭을 수행하며, 16x16, 32x32 탐색범위를 가지도록 구현하였다. 기존의 블록 기반 비디오 코덱에서는 탐색 방법으로 주로 full search, logarithmic search, three-step search 방법을 사용하였으며, 탐색 영역에서 동일한 블록을 찾는 방법으로 MSE(mean squared error), MAE(mean absolute error), SAE(sum of squared error) 등이 사용되었다. 제안한 비디오 코덱에서는 모바일 환경을 고려하였기 때문에 연산량이 적은 tree-step 탐색방식[22]과 식 (5)에 나타난 SAE를 사용한 블록 매칭을 수행하였다. C_{ij} 는 현재 프레임 내에서 움직임 추정이 수행되는 블록을 의미하며, P_{ij} 는 이전 프레임 내의 블록을 의미한다.

$$SAD = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (|C_{ij} - P_{ij}|) \quad (5)$$

픽셀의 변화가 많을 경우, 비슷한 블록을 찾기 어려우므로 이미지를 2배 확대한 후 블록 매칭을 수행하는 1/2pel 정밀도 움직임 벡터 추정을 적용하였다.

움직임 추정의 결과로서 생성되는 움직임 벡터는 가변길이 부호화로서 허프만 코딩 방법을 적용하여 프레임 복호를 위해 디코더로 전송하게 된다.

3.4.2 움직임 보상

움직임 보상은 이전 프레임과 움직임 벡터로서 예측 프레임(predicted frame)을 생성하는 과정이다. 예측 프레임은 이전 프레임으로부터 움직임 벡터만큼 매크로 블록을 이동시켜 생성되며, 예측 프레임과 현재 프레임간의 차로써 오차 프레임을 생성한다. 그림 13은 현재 프레임과 2번째 프레임간의 intercoding에 대하여 나타낸다. 그림 13의 (a)는 이전프레임에 해당하는 1번째 프레임과 움직임 벡터 값에 의해 생성된 예측 프레임을 나타낸다. 그림 13의 (b)는 2번째 프레임 즉, 현재 프레임을 나타낸다. 현재 프레임과 예측프레임간의 차를 구함으로서 그림 13의 (c)와 같이 오차 프레임(error frame)을 생성할 수 있다. 생성된 오차 프레임은 다시 중요 객체 영역과 비중요 객체 영역으로 분리되며, 3.1절에서 설명한 패딩 방법을 수행된 후 각 객체별 압축이 수행된다. 오차프레임으로부터 분리된 각 객체는 그림 13의 (d)와 (e)에 나타내었다.

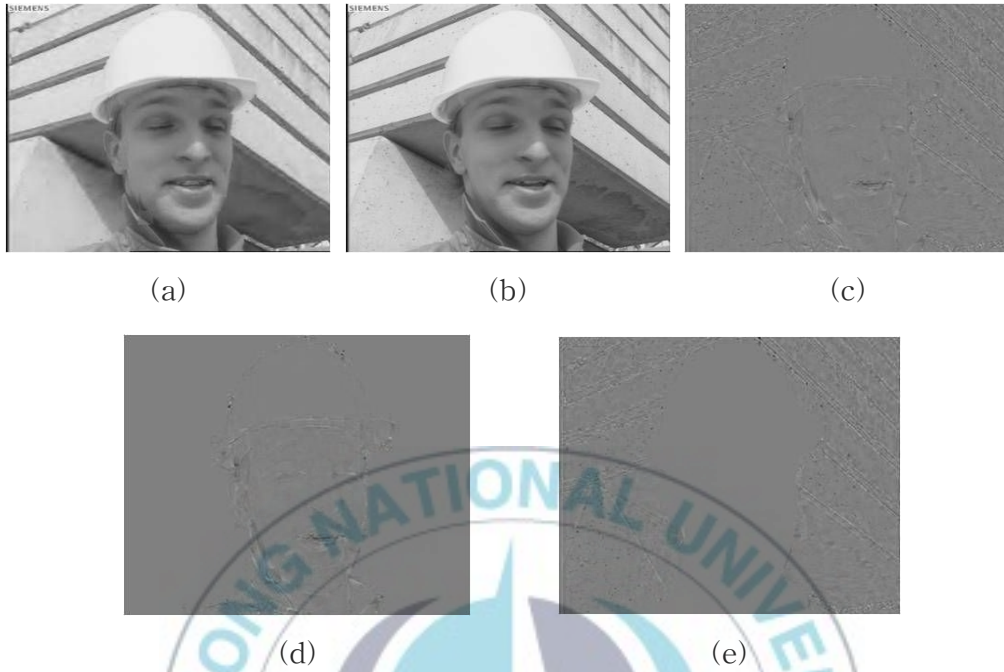


그림 13. 첫 번째 프레임과 두 번째 프레임 간의 intercoding 수행 예; (a) 예측 프레임, (b) 현재 프레임, (c) 오차 프레임, (d) 오차 프레임으로부터 분리된 중요 객체 및 (e) 비중요 객체

3.5 중요 객체에 대한 SNR 스케일러빌리티

중요 객체에 대하여 더욱 향상된 화질을 제공하기 위하여 그림 14와 같이 고위 계층(enhancement layer)의 생성과정을 블록도의 점선으로 나타내었다.

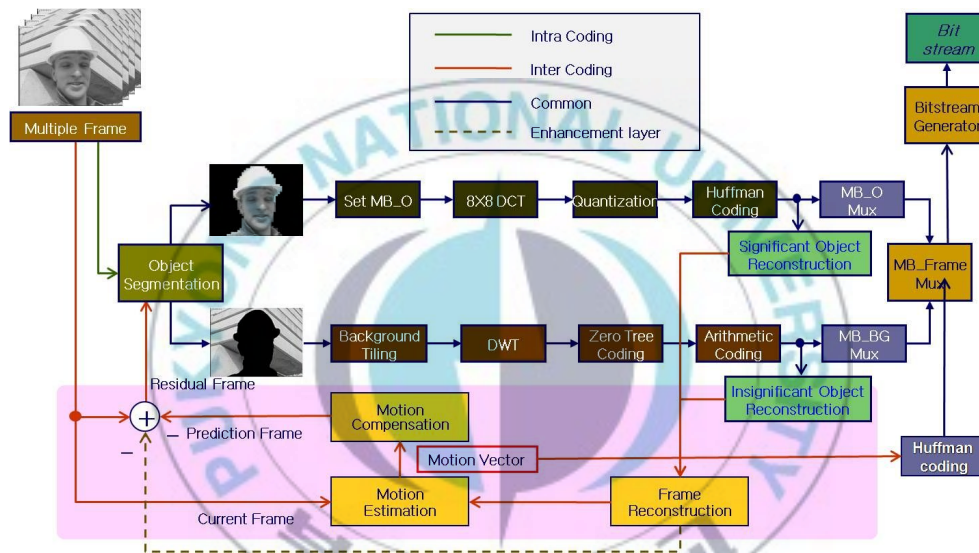


그림 14. SNR 스케일러빌리티가 추가된 중요 객체 부호화 블록도

중요 객체의 SNR 스케일러빌리티는 압축과정에서 양자화에 의한 손실된 데이터를 보상하는 방법이다. 원본의 중요 객체와 복원된 중요 객체의 차이를 구하여 잉여(residual) 프레임을 생성하며, 생성된 잉여 프레임을 압축하여 전송하게 된다. 그림 15에 잉여 프레임의 생성 예를 나타내었다.

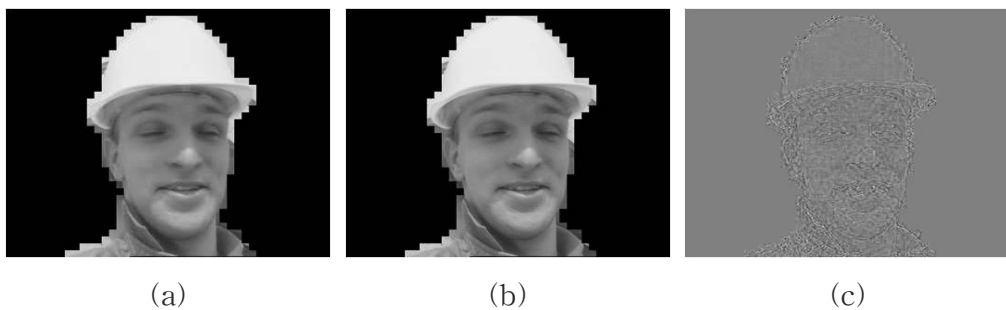


그림 15. 잉여 프레임 생성 예; (a) 원본 중요 객체, (b) 복원된 중요 객체 및 (c) 생성된 잉여 프레임

잉여 프레임에 대한 비트스트림은 디코더 단에서 복원되며, 복원된 비디오 시퀀스의 중요 객체에 손실된 화질을 보상해 줌으로써 중요 객체에 대한 더욱 향상된 화질을 제공하게 된다.



3.6 데이터 구조 설계

Decoder 단에서 부호화된 비트스트림의 복호를 위하여 데이터 구조를 정의하였다. 데이터 구조는 H.264 비디오 코덱에서 사용되고 있는 데이터 구조를 기반으로 설계하였다. H.264는 시작 단계부터 네트워크 전송을 목적으로 개발되었으며, 비디오 콘텐츠의 효율적 전송을 위해 NAL(network abstraction layer)로 구성된다. 제안하는 비디오 코덱 역시 네트워크 전송을 목적으로 하고 있으며, 객체에 독립적인 데이터 복호 및 향상 기법 적용을 위하여 NAL 구조를 기반으로 데이터 구조를 재정의하였다.

먼저 NAL에 대해 간단히 살펴보도록 한다. NAL은 그림 16과 같이 NAL unit과 페이로드로 이루어지며, 페이로드에는 RBSP(raw byte sequence payload)가 위치하게 된다. RBSP에는 SPS(sequence parameter set), PPS(picture parameter set)등의 정보를 알려주는 파라미터 셋과 VCL(video coding layer)에 해당하는 RBSP가 포함된다. VCL은 실제 H.264 압축된 데이터를 나타내며 VCL 데이터는 NAL unit를 통해 외부와 인터페이스가 이루어지게 된다. NAL unit은 4바이트 시작 패턴(start pattern)으로 시작하며 다음으로 1바이트 크기를 갖는 NAL header가 위치한다. NAL 헤더는 그림 17과 같이 forbidden_zero_bit, NAL_ref_idc, NAL_unit_type으로 구성된다. Forbidden_zero_bit는 항상 '0'으로 설정되어야 하며, NAL_ref_idc에 정의된 값은 NAL 헤더에 이어 나오는 페이로드가 inter 코딩 시 참조 픽처를 구성하는데 사용되는지의 여부를 확인한다. 다음에 위치하는 NAL_unit_type이 NAL 헤더 뒤에 위치하는 페이로드가 실제로 어떤 RBSP를 포함하는지를 나타내며, 실제 H.264의 NAL_unit_type을 표 1에 나타내었다.

Starter code 4Byte(0x00000001)	NAL header 1 Byte	Payload
--	-----------------------------	----------------

그림 16. H.264의 NAL 구조

Forbidden_zero 1bit(0)	Nal_ref_idc 2bit	Nal_unit_type 5bit
----------------------------------	----------------------------	------------------------------

그림 17. H.264의 NAL 헤더 구조

표 1. H.264 코덱의 NAL_unit_type 정의

NAL unit type	Contents of NAL unit and RBSP syntax structure
0	Unspecified
1	Coded slice of a non-IDR picture
2	Coded slice data partition A
3	Coded slice data partition B
4	Coded slice data partition C
5	Coded slice of a IDR picture
6	SEI(supplemental enhancement information)
7	SPS(sequence parameter set)
8	PPS(picture parameter set)
9	Access unit delimiter
10	End of sequence
11	End of stream
12	Filler data
13-23	Reserved
24-31	Unspecified

제안하는 비디오 코덱에 적합한 데이터 구조를 설계하기 위하여 복호화에 필수적인 정보들과 객체, 예측 모드, 스케일러블(scalable) 모드 등을 고려하여 NAL_ref_idc, NAL_unit_type을 표 2 및 표 3과 같이 재정의 하였다.

표 2. NAL 헤더 내 NAL_ref_idc 재정의

NAL_ref_idc	Contents of NAL_ref_idc
0	Only intra coding
1	Coded IDR object bitstream
2	Coded non IDR object bitstream
3	A content of the NAL unit is used to consist of reference picture.

표 3. NAL 헤더 내 NAL_unit_type 재정의

NAL_unit_type	Contents of NAL unit and RBSP syntax structure
0	Unspecified
1	IDR bitstream for significant object
2	IDR bitstream for insignificant object
3	Non IDR bitstream for significant object
4	IDR bitstream for insignificant object
5	Residual data for significant object in the SNR scalability mode
6	Decoding information for a sequence
7	Offset data of significant object
8	Motion vector information
9	End of sequence
10	End of stream
11-12	Reserved

Nal_ref_idc가 '0'으로 설정되면, 예측코딩이 적용되지 않고 모든 프레임에 대해 intra 코딩된 비디오 스트림이 페이로드에 위치함을 의미한다. '1'과 '2'는 예측 코딩이 적용된 intra/inter 코딩된 페이로드를 구분한다. '3'은 페이로드의 내용이 참조 픽처를 구성하는데 필요한 내용임을 의미하며, 복호화를 위한 각종 파라미터 셋과 중요 오브젝트의 오프셋 데이터, 움직임 벡터가 여기에 해당된다. Nal_unit_type에는 중요/비중요 객체의 intra/inter 코딩 여부, 복호화를 위한 파라미터 셋, 중요 객체의 오프셋 데이터, 그리고 움직임 벡터 정보 등을 구분하였다. 추후 추가되는 기능을 위해 '11-31'을 reserved로 설정하였다.

중요 객체에 대한 NAL은 intra/inter 코딩된 비트스트림을 구분하며, 그에 대한 NAL 구조 및 헤더 구성을 그림 18에 나타내었다.

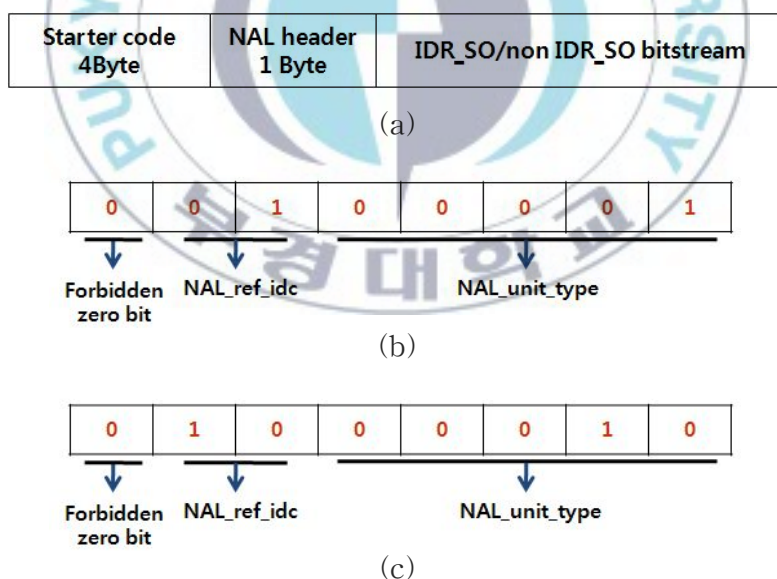


그림 18. IDR_SO/non IDR_SO NAL unit 구성; (a) SO NAL unit 구조, (b) IDR_SO NAL 헤더 및 (c) NonIDR_SO NAL 헤더

4바이트 크기의 시작 코드(starter code) 이후 페이로드의 형태를 정의하는 1바이트 크기의 NAL 헤더가 위치한다. 페이로드영역에 중요 객체의 비트스트림이 위치하며, 모든 페이로드의 최초 2바이트는 비트스트림의 크기를 저장한다. Intra 코딩된 중요 객체는 IDR_SO(instantaneous decoding refresh significant object) NAL로 정의하였으며, IDR_SO의 헤더를 그림 18의 (b)에 나타내었다. Inter 코딩된 중요 객체 대한 NAL을 Non IDR_SO로 정의 하였고, Non IDR_SO NAL 헤더를 그림 18의 (c)에 나타내었다.

비중요 객체에 대한 NAL 구조 및 헤더 구성은 그림 19에 나타내었다.

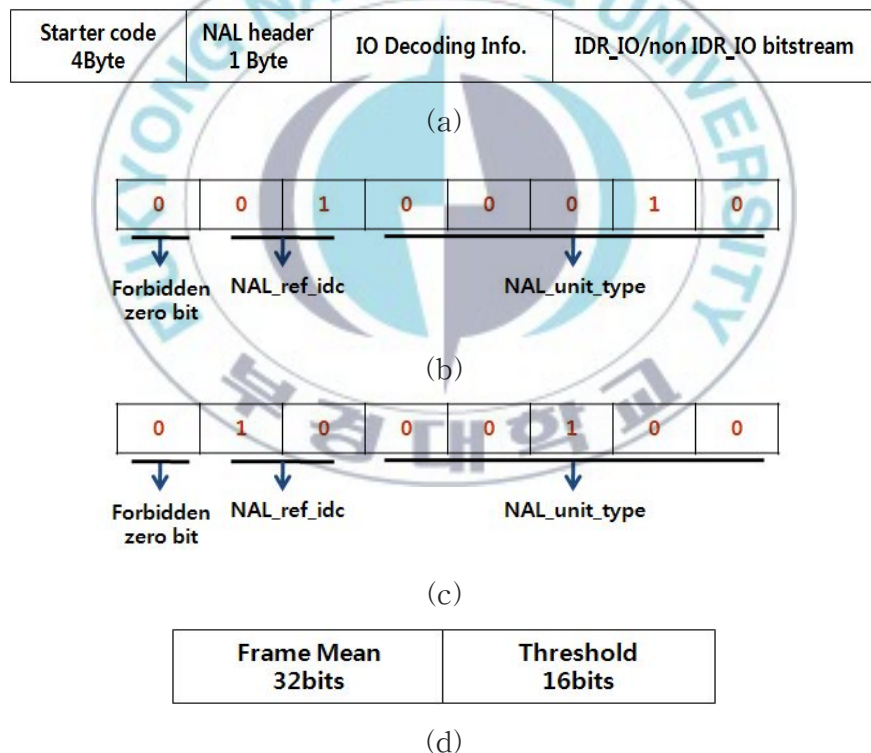


그림 19. IDR_IO/non IDR_IO NAL unit 구성; (a) IO NAL unit 구조, (b) IDR_IO NAL 헤더, (c) Non IDR_IO NAL 헤더 및 (d) IO Decoding Info.

비중요 객체의 NAL 구조는 그림 19의 (a)에 나타내었다. 비중요 객체의 페이로드는 복호를 위한 정보와 비중요 객체 비트스트림으로 구성된다. Intra 코딩된 비중요 객체는 IDR_IO(instantaneous decoding refresh insignificant object) NAL로 정의하였으며, IDR_IO의 헤더를 그림 19의 (b)에 나타내었다. Inter 코딩된 비중요 객체 대한 NAL을 Non IDR_IO로 정의 하였고, Non IDR_IO NAL 헤더를 그림 19의 (c)에 나타내었다. 비중요 객체는 프레임 단위로 웨이블릿 기반의 코딩이 수행되며, 비중요 객체의 복호를 위해 웨이블릿 기반 코딩 시 사용되는 프레임 평균 값(4바이트)과 임계값(2바이트)을 전송한다.

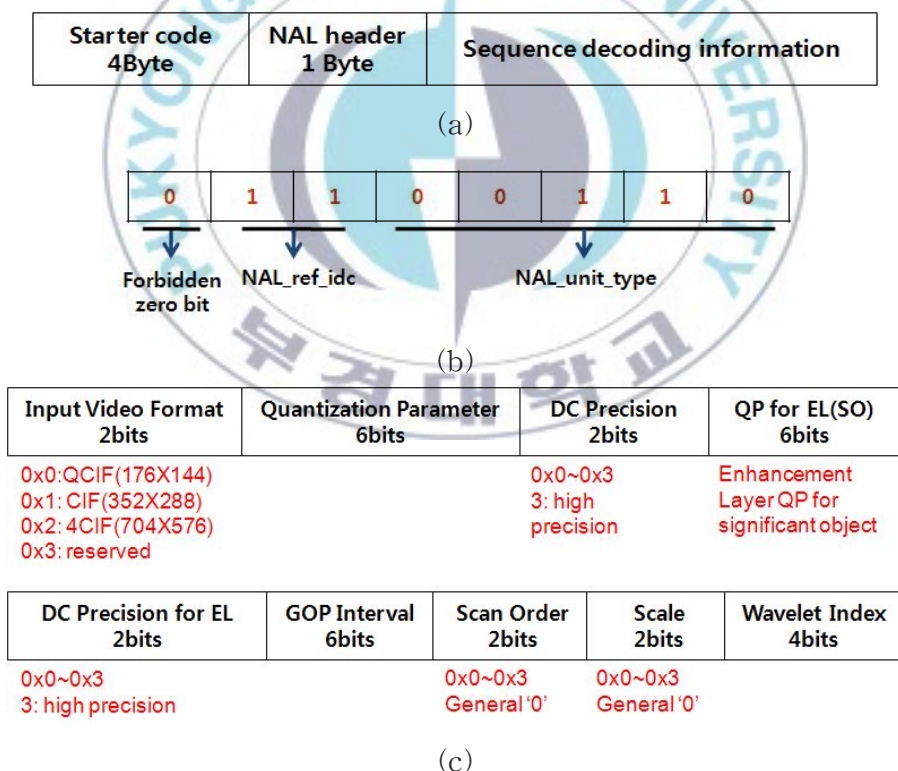


그림 20. DI NAL unit 구성; (a) DI NAL 구조, (b) DI NAL 헤더 및 (c) 데이터 비트 필드

그림 20의 (a)는 DI(decoding information) NAL 구조를 나타낸다. DI NAL의 NAL 헤더는 그림 20의 (b)에 나타내었다. 페이로드 영역에는 4바이트 크기를 갖는 복호화 정보들이 위치하며, 그 비트 필드(bit field)는 그림 20의 (c)에 나타내었다.

다음으로 SOOD(significant object offset data) NAL 구조를 살펴보도록 한다. 중요 객체에 대한 압축은 8x8 매크로 블록 단위로 수행된다. 압축될 매크로 블록은 프레임 내에 중요 객체가 포함된 블록만을 대상으로 하며, 디코더 단에서의 중요 객체의 복호화를 위하여 매 프레임마다 슬라이스 단위로 매크로 블록의 오프셋 포인트와 중요 객체 매크로 블록의 개수를 전송하게 된다. SOOD NAL의 NAL 헤더를 그림 21의 (a)에 나타내었다. 그림 21의 (b)는 오프셋 데이터의 구성 예를 보여준다.

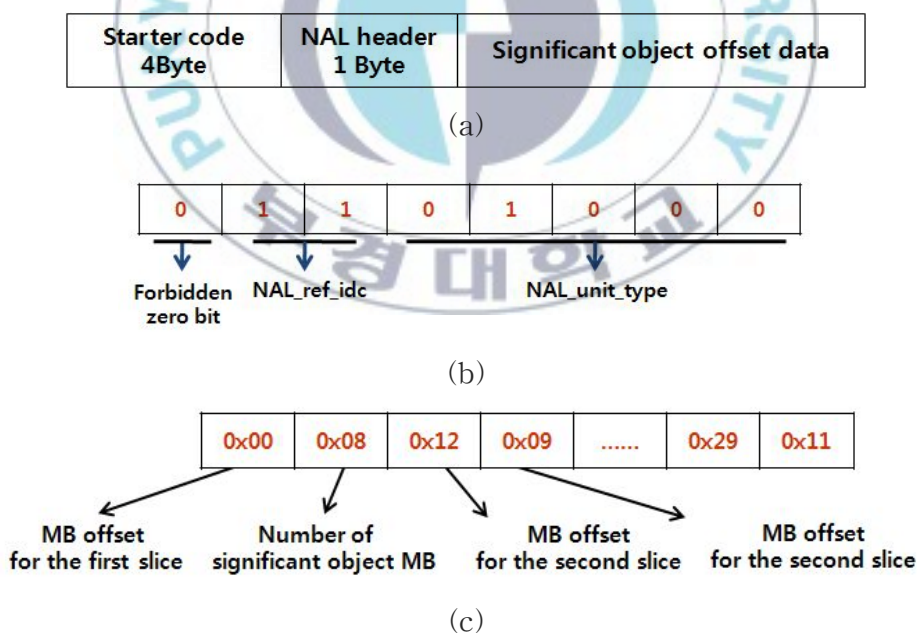
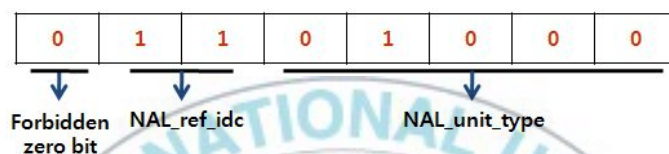


그림 21. SOOD NAL unit 구성; (a) SOOD NAL 구조, (b) SOOD NAL 헤더 및 (c) 오프셋 데이터 구성 예

MVI(motion vector information) NAL은 그림 22의 (a)에 나타낸 것과 같이 MVI NAL 헤더 다음으로 모션 벡터 값에 대해 허프만 코딩된 비트 스트림이 위치한다. MVI NAL 헤더는 그림 22의 (b)에 나타내었다.

Starter code 4Byte	NAL header 1 Byte	Motion vector information
-----------------------	----------------------	---------------------------

(a)



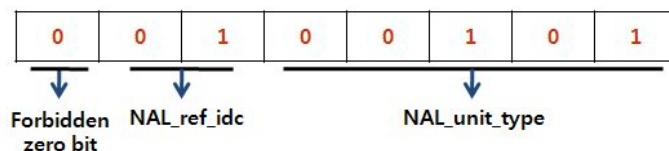
(b)

그림 22. MVI NAL unit 구성; (a) MVI NAL 구조 및 (b) MVI NAL 헤더

제안하는 코덱은 중요 객체에 대하여 더욱 향상된 화질을 제공하기 위한 객체 기반의 스케일러빌리티 기능을 제공한다. 이에 해당하는 NAL이 ELSO(enhancement layer for significant object) NAL이며, 그 구조와 NAL 헤더를 그림 23과 같이 정의 하였다.

Starter code 4Byte	NAL header 1 Byte	Residual data for significant object
-----------------------	----------------------	---

(a)



(b)

그림 23. ELSO NAL unit 구성; (a) ELSO NAL 구조 및 (b) ELSO NAL 헤더

IV. 실험 결과 및 분석

본 장에서는 기존의 비디오 코덱과 비교하여 제안한 비디오 코덱의 성능 평가 및 그 결과를 고찰 해본다. 기존 모바일 영상통화 환경에서 사용되고 있는 H.263과 성능을 비교하였으며, ITU-T TMN(test model near-term) 8 버전을 기반으로 하는 H.263+ 코더를 사용하였다. 실험은 입력 비디오의 휘도(luminance) 성분에만 한하여 수행하였으며, 제안한 방법으로 압축된 비디오와 H.263으로 압축된 비디오의 비트율을 같도록 하여 그 화질을 비교하도록 하였다. 화질 평가를 위해 원본 비디오와 복호화된 비디오 간의 PSNR(peak signal to noise ratio)을 사용하였다. PSNR은 식 (6)에 의해 구해진다.

$$PSNR [dB] = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \quad (6)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |O(i,j) - R(i,j)|^2 \quad (7)$$

실험 영상으로는 CIF(352×288) 해상도와 30fps의 프레임율을 갖는 Foreman 비디오를 사용하였다. Foreman 비디오 영상의 경우, 사람과 배경의 조합으로 프레임이 구성되어 있어 영상통화 상황으로 간주할 수 있다. 부호화 환경으로 GOP 구조를 '15'로 설정하였으며, 100 프레임에 대하여 비트율을 달리하여 객체별 화질을 비교하도록 하였다. 중요 객체에 대한 화질 변화를 비교를 위해 비중요 객체의 비트율을 intra, inter 프레임에 대

하여 각각 0.2bpp, 0.005bpp로 고정하여 실험하였다. 중요 객체의 비트율은 양자화 파라미터 값으로 조절된다. 표 4에 동일 비트율상에서 제안한 방법과 H.263으로 압축된 비디오에 대한 평균 PSNR을 비교하였으며, 그림 24부터 그림 27의 그래프에서는 비트율별 각 프레임에 대한 PSNR을 비교하였다.

표 4. 제안한 방법과 H.263 방법에 대한 평균 PSNR 비교

CODEC \ bpp(CR)		0.204 (1/39)	0.166 (1/48)	0.144 (1/55)	0.129 (1/61)
제안한 OBVC PSNR (dB)	SO bpp(QP)	0.40(12)	0.30(16)	0.25(20)	0.21(24)
	Frame	31.13	31.20	30.99	30.83
	SO	37.70	36.16	35.02	34.11
H.263 PSNR (dB)	Frame	34.33	33.76	32.82	32.60
	SO	34.78	34.02	33.37	33.01

※ OBVC - object based video coding, CR - compression ratio,
QP - quantization parameter, bpp - bits per pixel, SO - significant object

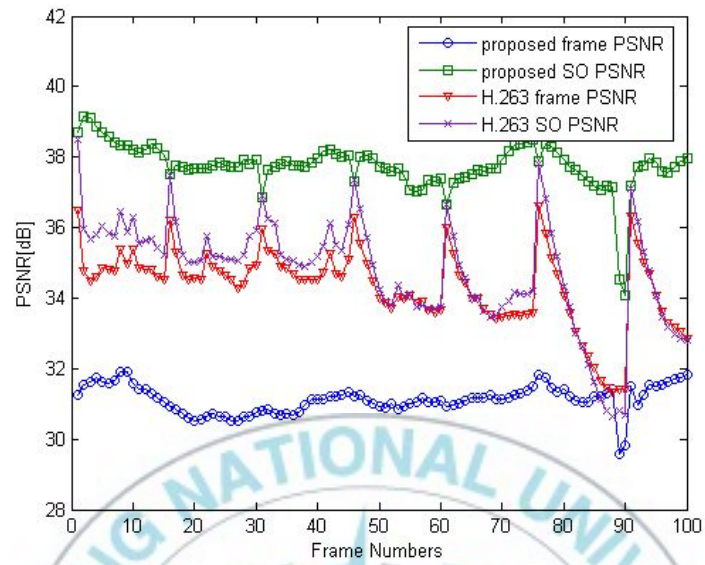


그림 24. 0.204 bpp에서 복호된 비디오의 프레임별 PSNR 비교

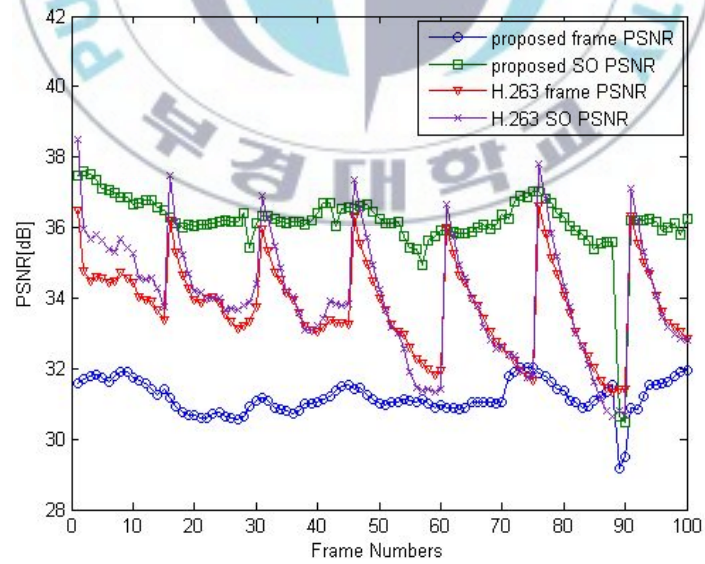


그림 25. 0.166 bpp에서 복호된 비디오의 프레임별 PSNR 비교

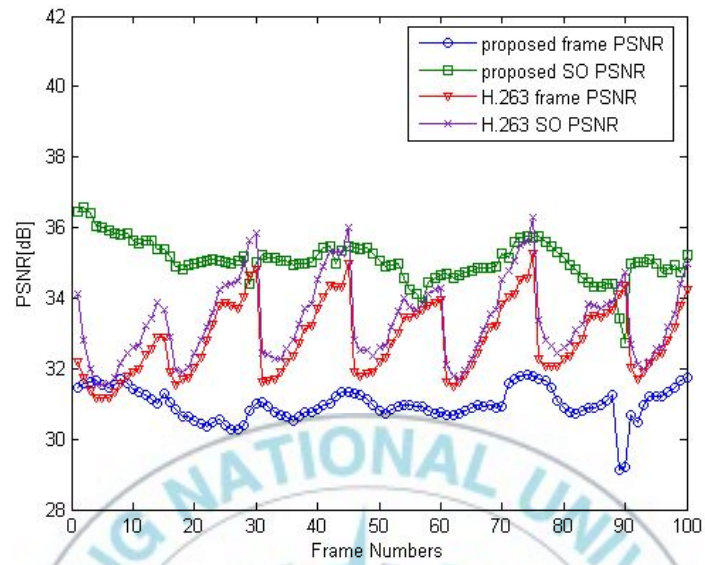


그림 26. 0.144 bpp에서 복호된 비디오의 프레임별 PSNR 비교

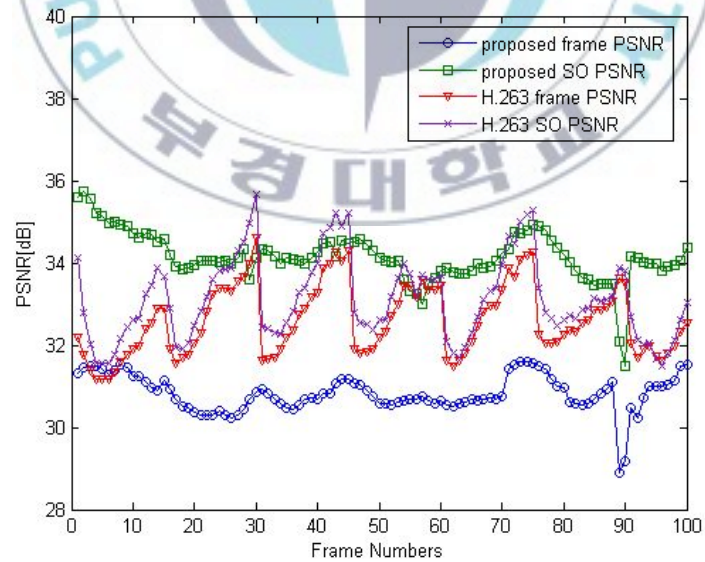


그림 27. 0.129 bpp에서 복호된 비디오의 프레임별 PSNR 비교

표 4는 100 프레임에 대하여 압축을 수행 후 복호한 프레임에 대한 평균 PSNR 값을 나타낸 것이다. 표 4의 결과에서 동일한 비트율상에서 복호된 프레임에 대한 평균 PSNR은 H.263 방법이 높게 나타나지만, 중요 객체에 대한 평균 PSNR은 H.263 방법으로 복호된 중요 객체의 PSNR보다 높게 나타남을 확인 할 수 있었다. 그림 24부터 그림 27에 나타낸 그래프를 보면 대체적으로 제안한 방법의 중요 객체에 대한 PSNR이 높게 나타나는 것을 확인할 수 있었다. H.263 방법의 intra 코딩된 프레임에 대한 중요 객체 PSNR이 제안한 방법보다 높게 나타날 때도 있지만 intra 코딩된 프레임과 inter 코딩된 프레임간의 화질 차이로 인하여 비디오 재생 시 화질 열화 성분이 가시적으로 드러났다. 90번째 프레임에서 급격한 움직임에 의해 화질이 떨어지는 것을 제외하면, 제안된 방법의 중요 객체에 대한 화질 변화가 크지 않음을 확인할 수 있었다.

제안한 방법과 H.263 방법으로 복호된 프레임에 대한 화질을 시각적으로 확인하기 위하여 그림 28부터 그림 31에 각 비트율별 복호된 프레임과 중요 객체의 임의 영역을 확대하여 나타내었다.



(a)

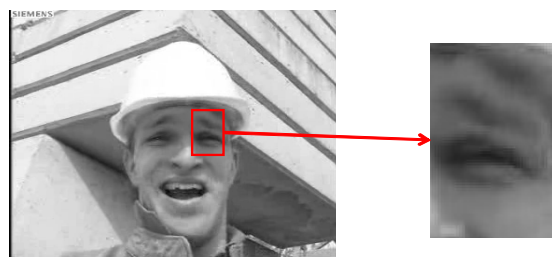


(b)

그림 28. 0.204 bpp에서 (a) 제안한 방법 및 (b) H.263으로 복호된 비디오의 3번째 프레임

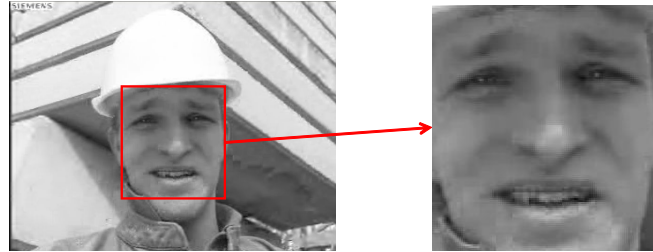


(a)

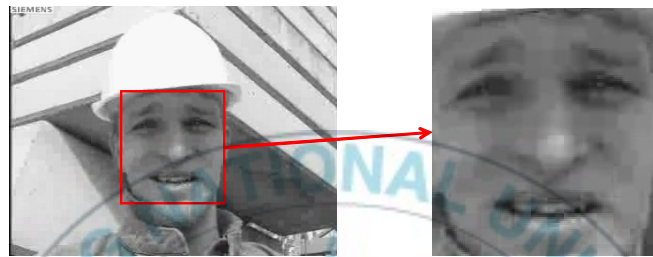


(b)

그림 29. 0.166 bpp에서 (a) 제안한 방법 및 (b) H.263으로 복호된 비디오의 25번째 프레임

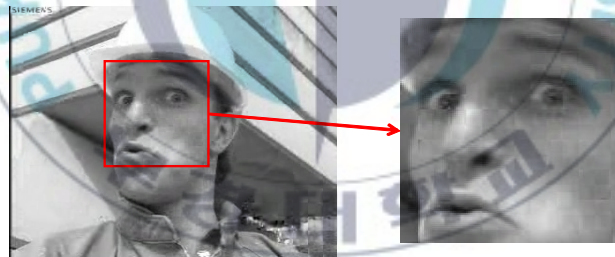


(a)

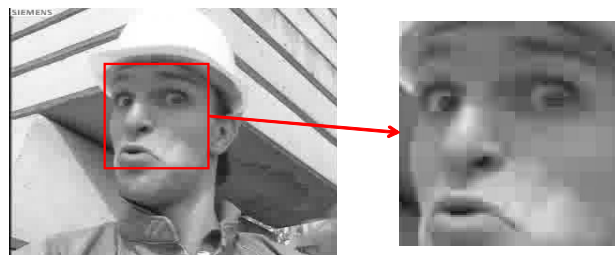


(b)

그림 30. 0.144 bpp에서 (a) 제안한 방법 및 (b) H.263으로 복호된 비디오의 50번째 프레임



(a)



(b)

그림 31. 0.129 bpp에서 (a) 제안한 방법 및 (b) H.263으로 복호된 비디오의 90번째 프레임

그림 28-31에 보이는 것처럼 동일한 비트율상에서 제안한 방법의 비중요 객체에 대한 화질은 고압축이 수행되어 H.263 방법의 비중요 객체 영역보다 화질이 좋지 않다. 하지만 높은 비트율을 부여하여 압축을 수행한 중요 객체 영역에서는 H.263보다 향상된 화질을 가진다. 그림 28, 30, 31에서 확대된 중요 객체 영역을 비교하면 H.263 방법의 중요 객체 영역에서 뚜렷한 블록화 현상을 볼 수 있었다. 또한, 그림 29의 (b)에서는 고압축을 수행하여 고주파 성분의 제거로 인해 H.263 방법의 확대된 중요 객체 영역에서 사람의 눈동자 부분이 제거된 것을 확인할 수 있었다. 하지만 그림 29의 (a)에는 눈동자 부분이 손실되지 않고 복호된 것을 확인할 수 있었다. 이와 같이 복호된 그림 28-31의 복호된 프레임을 비교함으로써 제안된 방법이 동일 비트율상에서 중요 객체영역에서 더 우수한 화질을 가지는 것을 시각적으로 확인할 수 있었다.

중요 객체에 대하여 더욱 향상된 화질의 제공하기 위한 SNR 스케일러빌리티 모드를 제공한다. 표 5에 SNR 스케일러빌리티 모드의 하위 계층과 고위 계층의 중요 객체 영역 PSNR과 H.263 방법의 중요 객체영역 PSNR을 비교하여 나타내었다. 그림 32에는 제안한 방법의 하위 계층과 고위 계층 그리고 h.263의 복호된 첫 번째 프레임을 비교하였다. 잉여 프레임에 해당하는 비트스트림이 추가되었기 때문에 전체 페이로드에 대한 비트율이 증가하였다. 따라서 동일 비트율상에서 H.263 방법의 중요 객체의 화질과 하위 계층(base layer)의 화질이 비슷하게 나타나는 것을 확인할 수 있었다. 이는 고위 계층을 압축하는데 많은 비트를 할당하였기 때문이며, 추후 잉여 프레임을 더욱 효과적으로 압축하는 방법에 대한 연구가 필요하다. 하위 계층에 복호된 잉여 데이터를 더하여 고위 계층을 생성하며, 생성된 고위 계층의 화질은 H.263보다 우수함을 확인할 수 있었다.

표 5. SNR 스케일러빌리티 모드에서 H.263과 중요 객체의 화질 비교

	제안한 OBSVC			H.263	
	BL		EL (SO)	Frame	SO
	Frame	SO			
bpp	0.23			0.23	
PSNR(dB)	30.99	35.02	37.18	34.89	34.93

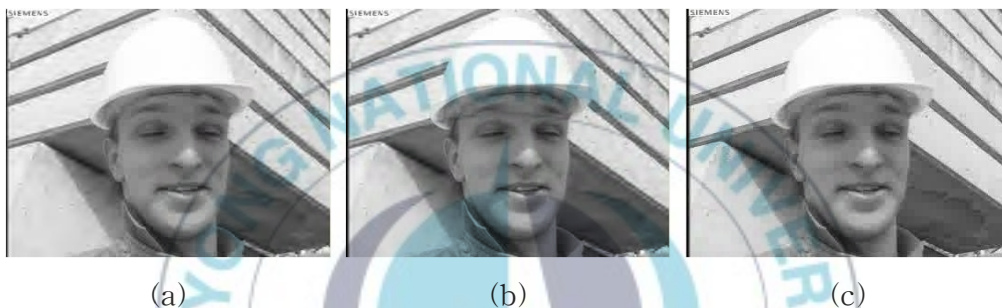


그림 32. SNR 스케일러빌리티 모드에서 첫 번째 프레임의 중요 객체 화질 비교를 위한 제안한 방법의 (a) 하위 계층과 (b) 고위 계층 및 (c) H.263 방법으로 복호된 프레임

본 절에서는 제안한 비디오 코딩 방법과 기존의 H.263 방법에 의한 성능을 비교하기 위하여 객관적인 화질 평가로서 PSNR을 비교하였으며, 시각적 화질 평가로서 복호된 프레임을 비교하였다. 동일 비트율상에서 제안한 방법의 중요 객체 영역에 대한 화질이 기존 H.263 방법의 중요 객체에 대한 화질보다 우수함을 확인할 수 있었다.

V. 결론

본 논문에서는 기존 3G 네트워크 환경에서 향상된 품질의 영상통화 서비스를 제공하기 위한 비디오 데이터 압축 방법을 제안하였다. 제안한 방법에서는 통화의 주체가 되는 사람의 얼굴 영역에 대한 화질 보장과 영상통화 시 발생하는 영상손실 성분의 최소화에 목표를 두었다. 이를 위해 비디오 데이터 압축 시, 프레임 내에서 중요 객체와 비중요 객체를 분리하였다. 얼굴 영역이 포함된 사람 영역을 중요 객체로 설정하였으며, 중요 객체에 대해서는 많은 비트를 할당하는 저압축을 수행하여 고화질을 갖도록 하였다. 반면, 배경영역에 해당하는 비중요 객체는 고압축을 수행하여 저화질을 갖도록 하는 압축 절차를 적용하였다. 중요 객체와 비중요 객체의 비트율은 trade-off 관계를 가지며, 이로써 기존의 제한된 대역폭에서 비트율을 만족하도록 하였다. 비중요 객체 영역에 대한 코딩의 경우, 기존 방법으로 고압축을 수행하면 블록 에러와 같은 영상손실이 발생하기 때문에 이를 해결하기 위해 DWT 기반의 프레임 단위 압축을 수행하였다. 또한, 중요 객체에 대해 더욱 향상된 화질을 제공하기 위한 SNR 스케일러빌리티 모델을 구현하였다.

실험 결과로부터 제안한 기법이 기존 영상통화 환경에서 사용되는 H.263에 비해 동일 비트율 상에서 중요 객체의 화질이 우수함을 확인할 수 있었다. 구체적으로 제안한 방법과 H.263의 비트율을 동일하게 두고 압축을 수행한 후에 복호된 프레임들의 평균 PSNR을 비교하였으며, 그 결과 제안한 방법의 평균 PSNR이 기존의 방법보다 높게 나타났다. 또한 주관적 화질 평가로서 복호된 각 프레임에 대한 시각적 평가에서도 제안한 방법의 중요 객체 영역이 우수한 화질을 가짐을 확인할 수 있었다.

본 논문은 기존 모바일 영상통화 환경의 제한된 대역폭에서 중요객체의 화질을 보장하는 객체기반의 비디오 코덱에 대하여 제안하였다. 현재까지의 모바일 환경을 고려한 객체기반 비디오 코덱의 연구는 아직 초기 단계로서, 제안한 방법의 성능 평가를 위하여 동일한 비트율상에서 각 객체에 대한 화질만을 비교하였다. 하지만 모바일 영상통화 환경을 고려하면 코덱의 실시간성 및 효율 또한 매우 중요한 요소이다. 따라서 향후 코드 최적화를 위한 연구가 진행되어야 하며, 객체기반의 움직임 추정 기법, 비중요 객체에 대한 패딩 기법 등 압축 효율을 높이기 위한 다양하고 지속적인 연구가 필요하다. 또한 중요객체에 대해 더욱 향상된 화질을 제공하기 위해 잉여 프레임에 대한 효율적인 압축 기법 역시 연구되어야 할 것이다.



참고 문헌

- [1] 오규태, “IMT-2000(CDMA2000과 WCDMA),” 방송과 기술 통권 143호, pp. 109-126, 2007년.
- [2] 박중현, “국내 HSDPA 서비스 수용의 영향요인 분석과 활성화 방향,” 전자통신동향분석 제 24권 제 1호, pp. 101-112, 2009년.
- [3] 권수갑, “HSDPA 개념과 동향,” 전자정보센터, 2005년.
- [4] 조경준, “HSDPA 서비스 동향,” 방송과 기술 통권 140호, pp. 112-126, 2007년 8월.
- [5] Video Codec for Audiovisual Services at p×64 kbit/s, ITU-T Rec. H.261, ITU-T, Version 1: Nov. 1990, Version 2: Mar. 1993.
- [6] Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s-Part 2: Video, ISO/IEC 11172-2(MPEG-1 Video), ISO/IEC JTC 1, Mar. 1993.
- [7] Generic Coding of Moving Pictures and Associated Audio Information-Part 2: Video, ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2 Video), ITU-T and ISO/IEC JTC 1, Nov. 1994.
- [8] Video Coding for Low Bit Rate communication, ITU-T Rec. H.263, ITU-T, Version 1: Nov. 1995, Version 2: Jan. 1998, Version 3: Nov. 2000.
- [9] Coding of audio-visual objects-Part 2: Visual, ISO/IEC 14492-2 (MPEG-4 Visual), ISO/IEC JTC 1, Version 1: Apr. 1999, Version 2: Feb. 2000, Version 3: May 2004.

- [10] Advanced Video Coding for Generic Audiovisual Services, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, Version 1: May 2003, Version 2: May 2004, Version 3: Mar. 2005, Version 4: Sept. 2005, Version 5 and Version 6: June 2006, Version 7: Apr. 2007.
- [11] Codec for circuit switched multimedia telephony service ; Modifications to H.324, 3GPP Technical Specification 26.111, Version 7.1.0: Jun. 2008, <http://www.3gpp.org/ftp/Specs/html-info/26111.htm>
- [12] Terminal for low bit-rate multimedia communication, ITU-T Rec. H.324, Spt. 2005, <http://www.itu.int/rec/T-REC-H.324/en>
- [13] Jabri. M.A. "The 3G-324M protocol for conversational video telephony," *IEEE MultiMedia*, vol. 11, no. 3, pp. 102-105, July 2004.
- [14] Packet-based multimedia communication systems. ITU-T Rec. H.323, 1998, <http://www.itu.int/rec/T-REC-H.323-200606-I/en/>
- [15] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding Using Wavelet Transform," *IEEE Trans. Image Processing*, pp. 205-220, Apr. 1992.
- [16] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Trans. Image Processing*, vol. 1. pp. 244-250, Apr. 1992.
- [17] J. M. Shapiro, "Embedded Image Coding Using Zerotree Wavelets Coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445-3462, Dec. 1993.

- [18] A. Said and W. A. Pearlman, "A New Fast and Efficient Image Codec based on Set Partitioning in Hierarchical Trees," *IEEE Trans. Circuits and System for Video Technology*, vol. 6, no. 3, pp. 243-250, Jun. 1996.
- [19] S. A. Martucci, I. Sodagar, T. Chiang, and Y. Zang, "A Zerotree Wavelet Video Coder," *IEEE Trans. Circuits and System for Video Technology*, vol. 7, no. 1, pp. 109 -118, July 1997.
- [20] A. S. Lewis and G. Knowles, "A 64 kB/s video Codec using the 2-D wavelet transform," *IEEE Computer Society Press, Data Compression Conf.*, pp. 196-201, Apr. 1991.
- [21] S. Cho and W. A. Pearlman, "A Full-Featured, Error-Resilient, Scalable Wavelet Video Codec Based on the Set Partitioning in Hierarchical Trees (SPIHT) Algorithm," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 157-171, Mar. 2002.
- [22] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 4, pp. 438-442, Aug. 1994.
- [23] S. H. Jeon, B. J. Jang, J. H. Kim, K. S. Moon, Y. S. Seo, S. H. Lee, and K. R. Kwon, "Video Watermarking Based on ROI Using Viola-Jones Face Detection," *APIS2009*, pp.79-82, Jan. 2009.
- [24] 전성혜, 장봉주, 서용수, 이석환, 권기룡, "Viola-Jones 얼굴 인식 방법을 이용한 ROI 기반 비디오 워터마킹," 멀티미디어학회 춘계학술대회논문집, pp. 218-221, 2008년 11월.

- [23] 전성혜, 장봉주, 서용수, 이석환, 권기룡, “모바일 영상통화 환경에 적합한 하이브리드 변환을 이용한 객체기반 비디오 CODEC 설계,” 대한전자공학회 하계학술발표대회논문집, pp. 989-990, 2009년 7월.
- [24] S. H. Jeon, B. J. Jang, Y. S. Seo, Y. D. Moon, S. H. Lee, K. R. Kwon, "Video CODEC Design Using Object-Based Hybrid Transform for the Mobile Video Telephony," *MITA2009*, pp. 113-114, Aug. 2009.
- [25] 전성혜, 장봉주, 서용수, 이석환, 권기룡, “모바일 영상통화 환경에 적합한 하이브리드 변환을 이용한 객체기반 비디오 CODEC 설계,” 정보처리학회 추계학술발표대회논문집, pp. 374-375, 2009년 11월.



감사의 글

굳은 의지와 부푼 마음을 안고 시작한 대학원 생활이 벌써 그 끝을 맺게 되었습니다. 아쉬움과 고마운 마음이 남아 2년이라는 시간이 너무도 짧게 느껴집니다. 저를 아껴주시고 도움을 주신 모든 분들께 감사 인사를 드리며, 이제 학교라는 울타리를 넘어 사회라는 더 큰 무대로 발을 들이려 합니다.

먼저 한없이 부족한 저에게 기회를 주시고, 아버지의 마음으로 지도 해주신 권기룡 교수님께 진심으로 감사드립니다. 교수님의 가르침은 평생을 살아가면서 되새기도록 하겠습니다. 참된 연구와 배움의 자세를 가르쳐 주신 이석환 교수님께도 진심으로 감사드립니다. 늘 격려와 걱정의 마음으로 저를 지켜봐주시며, 바른 길로 인도해 주신 서용수 교수님께도 항상 감사드립니다. 또한 저의 부족한 논문의 심사를 맡아 주신 정목동 교수님께도 감사드립니다.

대학원 생활을 하면서 가족, 친구보다 오랜 시간을 함께하며 정을 쌓은 MCSP 연구실 가족들 모두에게 감사의 마음을 전합니다. 연구실의 굳은 일을 도맡아 하며, 모두가 인정하는 연구실의 대들보 택영 형님, 대학원 내내 부족한 제게 가르침을 주기 위해 열의를 아끼지 않았던 봉주 형님 진심으로 감사하며, 죄송한 마음 전합니다. 다소 건방지고 무례한 후배를 귀엽게 봐주시고, 이끌어 주신 형님들 정말 잊지 못할 것입니다. 오랜 시간동안 선배로서 많은 도움을 준 혜정이, 항상 자신의 일처럼 저의 일을 도와주던 준희, 낯선 나라에서 공부를 하며 열심히 노력하는 XiaoZiao에게도 감사의 말을 전합니다. MCSP 연구실의 기둥으로 우뚝 설 대엽, 성범에게도 감사드립니다. MCSP를 졸업하신 모든 선배님들 진심으로 감사드립니다.

저를 믿고 지켜봐주신 저의 아버지, 어머니, 누나, 매형 진심으로 감사드리며, 모든 분들께 실망시키지 않도록 언제, 어디서든 노력하고 정진하는 모습 보여드리겠습니다.

2010년 2월 전 성 혜 드림