

통방융합 콘텐츠 서비스를 위한 스케일러블 비디오 부호화 및 적응 기술

□ 최해철, 강정원, 김재곤 / ETRI 디지털방송연구단

서 론

통신·방송의 융합은 각각 광대역화와 디지털화로 특징지어지는 통신과 방송 기술의 발전으로 디지털 콘텐츠 서비스의 새로운 패러다임으로 자리 잡아가고 있다. 통방융합 현상은 일반적으로 그림 1에서와 같이 다음 3 가지 유형으로 이야기 할 수 있다. 첫째 유형으로는 방송망과 통신망이 결합하는 네트워크 연동 및 통합, 둘째 통신 사업자의 방송 서비스 (예: IP-TV) 및 방송 사업자의 양방향 방송 서비스와 같은 사업자 및 서비스 융합, 마지막으로 통신 단말(휴대폰)에서 방송을 수신(예: DMB)하는 단말기 융합 등이 있다. 이러한 통방융합 환경에서 가장 대표적인 서비스로 모바일 단

말(스마트폰, PDA 등)에서 멀티미디어 방송 콘텐츠 서비스를 제공하는 모바일 방송이 부각되고 있으며, 콘텐츠 기술(description) 정보 및 사용자 정보를 활용하여 사용자가 원하는 콘텐츠를 사용자 소비환경(usage environment)에 맞추어 효과적으로 서비스하는 맞춤형 서비스가 개발되고 있다.

통방융합 환경에서 멀티미디어 콘텐츠를 서비스하기 위한 핵심 기술은 상호연동가능한(interoperable) 멀티미디어 프레임워크 기술과 통방융합 환경에 적합한 콘텐츠 생성 및 전송 기술을[1] 포함한다. 본 논문에서는 이중망과 다양한 단말로 구성된 통방융합 환경에서의 맞춤형의 콘텐츠 서비스를 제공하기 위한 주요 기술로 스케일러블(scalable) 비디오의 부호화와 적응 및 전송 기술의 동향을 다루고자 한다

스케일러블 비디오는 통방융합 환경에서의 다양한 시변의 망 특성 및 이중 단말 성능에 효과적으로 대응하기 위하여 다양한 계층(layer)의 스케일러빌리티(scalability)을 내재하고 있는 비디오 부호화 포맷(format)이다. 즉, 방송망과 인터넷망은 서로 다른 대역폭을 가지고 있을 뿐 아니라 한 종류의 망조차 그 대역폭은 시간에 따라 변화할 수 있다. 더욱이 다양한 소비 단말의 재생화면 크기와 연산 성능이 서로 다르기 때문에, 고정된 영상 포맷과 비트율(bitrate)로 부호화하고 전송하는 Non-scalable 비디오는 통방융합 환경에서 End-to-End QoS(Quality of Service)에 대한 사용자의 요구를 효과적으로 충족시킬 수 없다. 변환부호

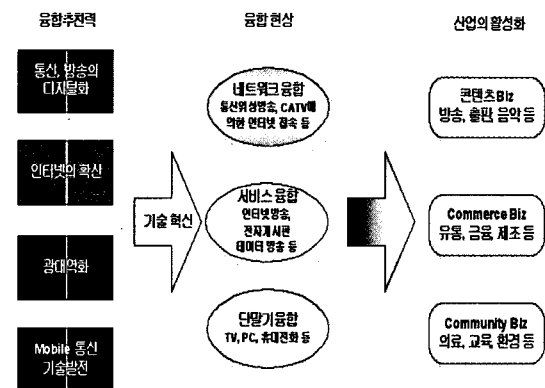


그림 1 통신과 방송의 융합현상

화(transcoding)나 비트량변환(transrating)를 통하여 제한된 범위에서의 영상포맷 및 비트율 적용이 가능하지만 실시간 적용이 요구되는 환경에서 비효율적이다.

이러한 콘텐츠 전달 및 소비 환경에 효과적으로 대응하기 위하여 비디오를 다양한 공간적 해상도와 화질, 시간적 해상도(즉, frame-rate)를 갖는 다수의 계층들을 하나의 비트열(bitstream)로 부호화하여, 접속망 상태에 적합한 비트율을 갖는 계층으로 전송하며 단말 성능을 만족하는 영상 포맷 및 화질의 비트스트림만 복호화 한다. 이렇게 부호화시에 미리 시간/공간/화질(temporal/spatial/quality)에 대하여 단계적인 자유도를 제공하는 것을 스케일러블 비디오 부호화(SVC: Scalable Video Coding) 기법이라 한다. 이 SVC는 단일 부호화 스트림으로 다양한 제한 환경에 끊임 없이 QoS가 보장된 비디오의 전송/소비를 제공할 수 있다.

기존의 MPEG-2 와 MPEG-4(part2) 비디오에서 계층 부호화에 기반한 스케일러블리티의 지원을 위한 시도가 있었지만, 이 표준들에서 정의된 비디오 스케일러블리티의 성능 수준은 매우 미흡하다. 현재 ISO/IEC MPEG(Moving Picture Experts Group)과 ITU-T VCEG(Video Coding Experts Group)의 JVT(Joint Video Team)에서 새로운 스케일러블 비디오 부호화 기법들을 가장 개선되고 발전된 기술들로서 표준화 하는 작업이 활발히 진행 중이며, 현재 WD 2와 JSVM(Joint Scalable Video Model) 2가 나왔고 2006년 7월 표준화 완료를 목표로 한다.

본 고에서는 통방융합 환경에서 스케일러블 비디오의 역할에 대해 설명하고, JVT에서 국제 표준화 중인 SVC 부호화 개요와 제공되는 스케일러블 기능들에 대해 살펴본다. 그리고 스케일러블 비디오가 이중망과 다양한 단말 환경에서 어떻게 효율적인 전송 및 적용을 제공할 수 있는지를 살펴본다.

스케일러블 비디오 부호화

스케일러블 비디오 부호화의 표준화 동향

MPEG에서는 MPEG-21 (part 13)을 시작점으로 SVC라는 소그룹(AdHoc)을 결성하였다. 2004년의

CFP(Call for Proposal)에서 다음 두 가지 범위의 시나리오를 설정하고 제안 기술을 검증 하였다. 첫 번째 시나리오는 넓은 범위의 스케일러블리티로 감시시스템, 방송, 저장장치를 적용대상으로 삼고 있으며, 1.2초 범위의 임의 접근(random access)을 요구한다. 두 번째 시나리오는 보다 제한된 범위의 스케일러블리티로 스트리밍(streaming), 이동통신 등의 분야에 관한 것이며 임의 접근의 제한은 없다.

기본적인 SVC의 프레임워크로 다음 2 가지의 기법이 제안 검증 되었다.

* DCT 기반의 기법: 기존 MPEG 표준들과 유사한 접근 방식을 취하며 이들과 호환성을 갖는다.

* 웨이블릿(wavelet) 기반 기법: 최초로 비디오 압축 도메인에서 재귀하지 않는(non-recursive) 구조를 제공한다.

검증 결과, Wavelet 기반 기법이 근본적으로 스케일러블리티를 포함하고는 있으나, wavelet 필터의 탭(tap) 길이가 영상 macroblock의 화소 수보다 길기 때문에 구조적으로 intra 모드를 제공할 수 없고, 이에 따라 압축 효율성 측면에서 HHI에서 제안한 DCT 기반의 기법보다 낮은 성능을 보여주었다. 따라서 DCT 기반 기법이 SVC의 기본 프레임워크로 채택 되었으며, 이를 기반으로 많은 알고리즘들이 연구, 개발되고 있다.

지난 2005년 1월 MPEG 홍콩 회의에서, 또 다른 비디오 국제 표준화 단체인 ITU-T의 VCEG이 동참하면서 JVT에서 스케일러블 비디오 표준화가 수행 중이며, 명칭은 MPEG-4 part 10 AVC AMD 1 혹은 H.264.4 이며 현재 WD(Working Draft) 2 와 J-SVM(Joint-Scalable Video Model) 2가 발표되었다.

스케일러블 비디오 부호화 개요

통방융합 환경에 대처하기 위한 가장 중요한 기술 중의 하나는 멀티미디어 콘텐츠의 중심 요소인 비디오를 생성/전송/소비함에 있어서 스케일러블리티를 제공하는 것이다. SVC는 여러 개의 비디오 계층을 하나의 비트열로 부호화하며, 각 층은 각각의 비트율, 프레임율, 영상 크기 및 화질을 가지고 있다. 즉 비디오를 공간적/시간적/화질적 차원의 임의 값을 가지는 비트

열로 부호화하며, 그 세 가지 차원의 조합에 따라 폭넓은 스케일러빌리티를 제공할 수 있다.

기술적인 관점에서, 하나의 스케일러블 비트열은 두 개 혹은 그 이상의 의존적인 계층으로 구성될 수 있다. 이 경우, 스케일러블 코덱(codec)은 하나의 기본계층(base layer)과 스케일러블 상위계층(enhancement layer)들로 구성된다. 여기서 기본계층 및 연속되는 상위계층의 정보가 함께 이용되어 보다 개선된 비디오 비트열을 만든다. 그 예로 화질 스케일러빌리티는 하나의 비트열로부터 동일한 공간 및 시간 해상도(dimension)를 갖지만 각각 다른 화질을 갖는 비디오 계층들을 만들어 낸다. 일반적으로 기본계층은 기본적인 비디오 화질을 제공하고, 연속된 상위계층은 이전 계층들로 만들어진 비디오보다 높은 화질을 갖도록 부호화한다. 마찬가지로 시간 및 공간 해상도에서도 동일한 원리를 적용하여 스케일러빌리티를 지원한다. 하나의 차원에 대한 스케일러빌리티로는 최적의 가변성과 적응성을 제공할 수 없는 응용분야가 있을 수 있으며, 이는 그림 2과 같이 시간, 공간, 화질 측면 모두의 스케일러빌리티를 필요로 한다.

그림 2은 세 가지의 스케일러블 기능을 지원하는 예이다. QCIF(176x144) 크기를 갖는 공간적 기본계층은 AVC 표준을 이용하여 15 Hz의 프레임율로 부호화된

다. 여기서 시간적 스케일러빌리티를 지원하기 위해 hierarchical B pictures 기법[2][4]이 사용된다. CIF(362x288) 해상도를 가지는 공간적 상위 계층은 30 Hz로 부호화되며, MCTF(Motion Compensated Temporal Filtering) 기법을 이용하여 4가지(15Hz, 7.5Hz, 3.75Hz, 1.875Hz) 계층의 시간적 스케일러빌리티를 제공한다. 각 공간 계층(QCIF, CIF)은 FGS(Fine Grain Scalability) 기법으로써 QCIF인 경우 41~80kbits/s, CIF인 경우 115~256kbit/s의 SNR 스케일러빌리티를 지원한다.

기존의 표준 비디오 부호화 기법들에 비해 SVC는 하나의 비트열에 스케일러빌리티를 갖는 계층들을 포함하는 것과 MCTF를 이용하는 것이 가장 큰 기술적 차이점이라고 할 수 있다. 다음 두 절에서 이 두 가지 특징적인 기술들에 대해 각각 기술한다.

스케일러블 부호기

SVC의 부호기에 대해 기술하고 있는 J-SVM은 다음 두 가지 방법을 통해 스케일러블 기능을 제공한다. 한 가지는 비트플레인(bit-plane) 산술코딩(arithmetic coding)과 같은 부호화 방법이고, 다른 하나는 기존의 여러 스케일러블 표준에서 사용되었던 계층기반 접근 기법이다. J-SVM에서는 전체 시간적/공간적/화질적

스케일러빌리티를 제공하기 위해 상기 두 가지 방법을 모두 사용하였다. 시간적 스케일러빌리티는 MCTF (Motion Compensated Temporal Filtering)로써 지원하고, 공간적인 스케일러빌리티는 down/up 표본화(sampling)를 통해 여러 해상도를 가지는 영상들을 구성함으로써 가능하게 하며, 화질(quality 혹은 SNR) 스케일러빌리티를 위해서는 비트플레인 부호화를 이용한다.

J-SVM 부호기 구조는 제공하

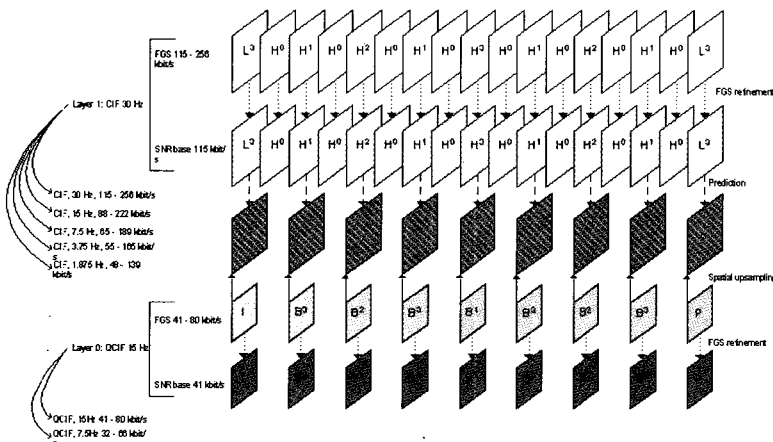


그림 2 SVC의 시간/공간/화질 스케일러빌리티의 예

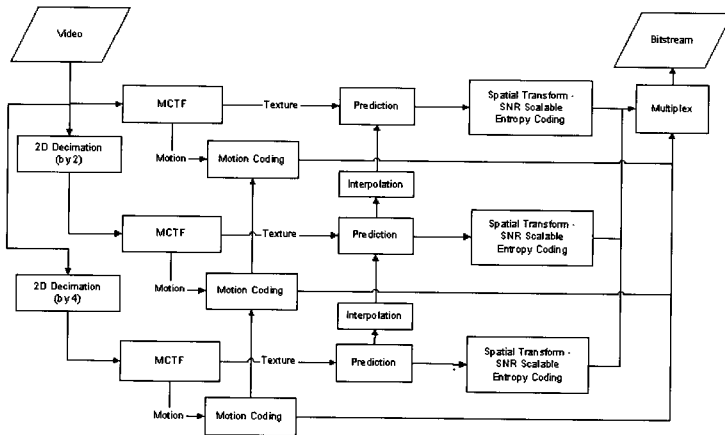


그림 3 공간적 3 계층의 스케일러빌리티를 제공하는 스케일러를 부호기

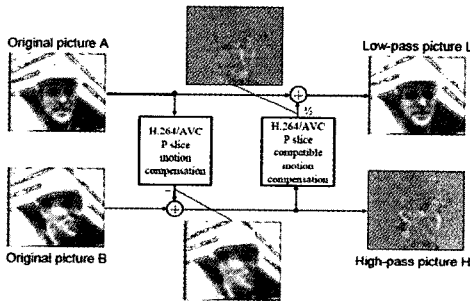


그림 4 Haar 필터를 이용한 MCTF 개요

려는 스케일러빌리티의 종류 및 단계에 따라 달라진다. 공간적 계층 및 MCTF를 이용한 부호기 구조는 그림 3과 같다. 공간적 스케일러빌리티를 위해 입력 영상의 공간적 해상도를 줄여 크기가 작아진 여러 개의 영상들을 구성하며 이때 MPEG-4의 down-sampling 필터를 사용한다. 그 후 공간적으로 분할된 각 계층의 입력 비디오 신호에 MCTF를 적용하여 시간적 스케일러빌리티를 지원한다. 이 과정의 결과는 시간적인 저주파 영상(L)과 고주파 영상(H)의 집합, 잔여 텍스처 정보, 블록 단위의 움직임 정보이다. 움직임 정보와 잔여 텍스처 정보는 AVC

알고리즘을 기초로 부호화하면서, 부가적으로 화질 스케일러빌리티를 제공하기 위해 CGS(Coarse Grain Scalability) 혹은 FGS(Fine Grain Scalability) 기법을 적용한다. 또한 전체적으로 각 공간 계층 간의 정보 중복성을 줄이기 위해, 움직임 정보를 그 자체로 부호화하지 않고 하위 계층에 있는 대응 블록의 움직임 정보와의 차이 혹은 그 수정 값을 포함시켜 부호화하는 방식을 이용할 수

있다. 잔여 텍스처(residual texture) 정보 및 intra macroblock의 예측 과정에 대해서도 동일한 개념의 방식을 AVC에서의 방식들과 함께 적용한다.

MCTF(Motion Compensated Temporal Filtering)

SVC에서 시간적 스케일러빌리티는 MCTF로써 지원한다. MCTF는 개념적으로 그림 4와 같다. 시간적으로 이웃하는 두 프레임의 원영상 A, B가 있을 때, 움직임 보상을 통하여 A 영상을 B 영상과 유사하도록 만든다. 그 결과 영상과 B 영상 간의 차이는 고주파 영상(H)가 되며, 그 차이의 반 값을 A 영상에 더해 주면 결국 $(A+B)/2$ 가 되므로, A와 B 영상의 평균 영상이 출력된다. 이를 저주파 영상(L)이라고 한다. 이와 같이 두 프레임의 원영상을 움직임 정보를 이용하여 시간적으로

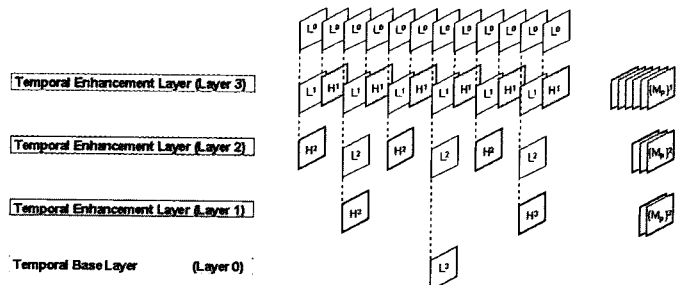


그림 5 3 분해(decomposition) 과정의 MCTF

고주파 영상과 저주파 영상으로 필터링하므로 이 기법을 MCTF라고 부른다. 실질적으로 GOP(group of picture) 단위의 영상열에 MCTF를 가하여 저주파 영상 셋과 고주파 영상 셋을 얻은 후, 저주파 영상 셋에 대해서 MCTF를 반복함으로써 그림 5와 같은 계층들을 구성하여 시간적인 스케일러빌리티를 지원한다.

그림 5의 경우 12 프레임의 영상으로 이루어진 GOP에 대해 세 번의 MCTF를 적용했으며, 이에 따라 저주파 영상 L3는 12 프레임의 평균값을 나타내는 시간적 기본 계층이 된다. 최소 프레임율을 요구할 때, 이 L3만을 전송함으로써 원영상열보다 1/12 낮은 프레임율을 지원할 수 있다. 그림의 고주파 영상 (H3)은 첫 번째 상위계층이 되며, 기본 계층의 L3와 함께 복호하여 (L2)를 만들어 낼 수 있다. 따라서 시간적 기본 계층 L3와 첫 번째 시간적 상위 계층 (H3)들과 이에 해당하는 움직임 정보 (Mp3)를 전송함으로써 수신단은 3 장의 L2 영상을 화면에 출력할 수 있다. 즉 원영상열과 비교 시 1/4의 프레임율을 지원하는 것이다. 이와 같은 방식으로 SVC는 시간적 스케일러빌리티를 지원한다.

MPEG-21 DIA를 이용한 비디오 적응 변환

비디오 콘텐츠를 이종망 특성 및 다양한 단말 성능에 적응적으로 끊임없이 서비스하기 위해서는 비디오 콘텐츠를 주어진 제한 조건을 만족하도록 변환하는 적응(adaptation)은 필수적으로 요구된다. 특히 3장에서 설명한 스케일러블 부호화기를 이용하여 부호화된 스케일러블 비디오는 이러한 비디오 적응을 고려한 것으로 부호화 방식으로, 주어진 네트워크 및 단말의 환경에 적합한 보다 효율적인 비디오 콘텐츠의 적응변환을 가능하게 한다는 특징을 갖는다. 그러나 비디오 콘텐츠가 네트워크/단말의 환경에 적응적으로 변환하기 위해서는 주어진 네트워크 및 단말의 환경에서 최적의 적응변환 값(즉, 비디오 콘텐츠에서 시간/공간/화질의 최적의 조합)을 결정하고, 그 값을 입력 받아 효율적인 적응변환을 수행하는 틀이 제공되어야만 한다.

MPEG-21 에서는 위에서 언급한, 비디오와 같은

디지털 아이템(DI, Digital Item)의 적응을 위한 표준 틀(기술자, Description) 개발을 목적으로 DIA(Digital Item Adaptation) 표준화 작업이 활발하게 진행되어 왔다. 현재 FDIS[15]에서는 Resource Adaptation Tool로 네트워크와 단말의 QoS를 위한 AdaptationQoS Tool이 채택되어 있다. 특히, 최적의 적응을 위하여 리소스(예, 비디오)의 다양한 제한, 이를 만족하기 위해 이용 가능한 적응 방법, 그리고 각 적응에 대한 유틸리티(utility)의 관계를 기술하기 위한 틀인 UtilityFunction이 현재 AdpatationQoS의 모듈로 포함되어 있다. 또한 리소스 비트스트림의 선택스를 상위레벨에서 기술함으로써 적응 엔진이 비트열 선택스를 이해하지 않고 비트열 포맷에 독립적으로 효율적인 적응을 하기 위한 BSD(Bitstream Syntax Description) Tool로 BSD/gBSD(generic BSD)들이 채택되어 있다. 이들 AdpatationQoS와 BSD/gBSD Tool은 독립적으로 사용될 수 있지만 복합적으로 사용할 경우 QoS를 고려한 효율적인 리소스 적응을 구현할 수 있다. AdaptationQoS는 소비환경(usage environment)의 제한 조건을 만족하는 Adaptation Operator를 QoS를 고려하여 (최적의) 결정, 선택을 가능하게 하고, BSD/gBSD는 결정된 Adaptation Operator에 따라서 비트열 선택스를 파싱하지 않고도 효율적인 적응을 가능하게 한다. 또한 AdaptationQoS와 BSD/ gBSD의 효과적인 결합을 위하여 BSDLink 틀이 제안 채택되어 있다. 아래 그림은 위에서 기술한 MPEG-21 DIA 기반의 비디오 적응 틀

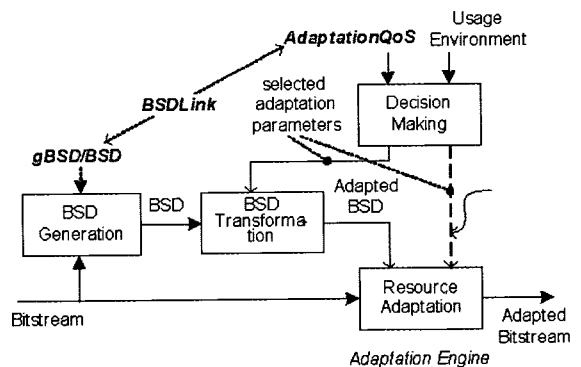


그림 6 MPEG-21 DIA 기반의 비디오 적응 틀 구성도

의 구성도를 나타낸 것이다.

결 · 론

현재 방송과 통신의 융합이 진행됨에 따라, 서로 다른 특성을 갖는 이종망과 다양한 재생화면 크기, 연산 성능을 가지는 단말기들이 혼재해 있다. 이러한 통방융합 환경에서 비디오 콘텐츠를 이종망을 통하여 다양한 단말에서 시간과 장소에 제한 받지 않고 End-to-End QoS가 보장된 품질로 단절 없이 서비스하기 위해서는, 비디오 콘텐츠에 대해 생성 단계에서부터 비디오에 대한 화질, 크기, 시간에 대한 스케일러빌리티를 지원해야 하며, 가능하다면 부호기와 복호기에서의 복잡도에 대한 스케일러빌리티를 포함할 수 있어야 한다.

따라서 본 고에서는 공간적/시간적/화질적 스케일러빌리티를 제공하는 SVC를 소개하고, MPEG과 ITU-T의 JVT에서 진행 중인 이의 국제 표준화 동향과 SVC 부호기의 기술적인 구조, 스케일러블 기능의 지원 방식에 대해 살펴 보았다. 더불어 효율적인 비디오의 적용 변환을 위해 MPEG-21 DIA를 이용한 적용 기술에 대해 설명 하였다.

결론적으로 SVC는 통방융합 환경에서 발생하는 사용자의 선호도, 소비 단말기와 네트워크의 성능 등 다양한 전달 및 소비 환경에 효과적으로 대처할 수 있으며, 스케일러빌리티를 갖는 계층들의 조합을 하나의 비트열로 부호화하여 적응적인 전달/소비를 허용함으로써 여러 소비 환경 각각에 차별적인 콘텐츠를 제공할 수 있다. 더욱이 동일한 생산 비용이 여러 개의 전달/소비 환경에 공유될 수 있기 때문에 비용과 효율의 생산성 측면에서도 SVC는 매우 매력적이다. 따라서 스케일러블 비디오는 통방융합 환경에서 보다 향상된 비디오 서비스와 새로운 시장의 기회를 제공하며, 이 경향은 통방융합 추세를 더욱 촉진시킬 것이다

[참고 문헌]

[1] 김태희, 김재곤, 홍진우, "통신(방송 융합 콘텐츠 서비스 기술: 멀티미디어 프레임워크," 한국통신학회지, Vol. 22, No. 4, pp. 37~48, 2005년 4월

- [2] ITU-T and ISO/IEC JTC1, "Advanced Video Coding for Generic Audiovisual Services," ITU-T Recommendation H.264 - ISO/IEC 14496-10 AVC, 2003년.
- [3] ITU-T and ISO/IEC JTC1, "Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods", JVT-N046, 2005년 1월.
- [4] ITU-T and ISO/IEC JTC1, "JSVM 2 Software", JVT-O203, 2005년 4월.
- [5] ITU-T and ISO/IEC JTC1, "Scalable Video Coding - Working Draft 2", JVT-O201, 2005년 4월.
- [6] 김재곤, 김규현, IT 신성장 동력: 통신·방송 융합을 고려한 맞춤형 방송 기술 동향, 주간기술동향, 통권 1168호, 2004년 10월.
- [7] 강정원, 이회경, 김재곤, 맞춤형 콘텐츠 서비스를 위한 TV-Anytime 기반 콘텐츠 패키징, 방송공학 회논문지, 제9권, 제4호, 2004년.
- [8] 박구만, "방송통신 융합 네트워크의 발전 방향 분석," 방송공학회지 제8권 4호, 2003년 12월.
- [9] 변상규, "방송통신 융합을 주도하는 DMB 서비스" ETRI CEO Info 3호, 2004년 5월.
- [10] 김용배, "무선인터넷서비스의 대중화를 선도하는 휴대 인터넷 기술" ETRI CEO Info 6호, 2004년 6월.
- [11] 김성민, 지상파 DMB 시장현황 및 전망, TTA저널, 제94호, pp.39-46, 2004년.
- [12] ROA Group, 모바일 콘텐츠 시장의 향후 트렌드 변화와 전망, ROA-00-2-027, 2004년.
- [13] 김재곤, 최진수, 김진웅, 맞춤형방송 기술과 표준화 동향, 전자통신동향분석, 제19권, 제4호, 2004년 8월.
- [14] ENTHRONED Deliverable D10, "Scalable Audio and Video Codec," Contract No. 507637, 2004년 11월.
- [15] ISO/IEC 21000-7: 2004, Information Technology - Multimedia Framework - Part 7: Digital Item Adaptation.