주 제

방송통신 융합서비스를 위한 콘텐츠 적응 기술

ICU 노용만, 정용주, Truong Cong Thang

차례

요 약

방송과 통신이 융합하는 새로운 환경의 도래에 따라 사용자에게 불편함 없이 언제 어디서나 멀티미디어 콘텐츠를 접근(universal multimedia access)하게 할 수 있는 기술적 필요성이 대두되고 있다. 본 논문에서는 이러한 기술들 중에서 가장 중요한 위치를 차지하고 있는 콘텐츠 적응(content adaptation)에 대해 논한다. 특히 현재의 MPEG-21 표준에서의 콘텐츠 적응에 대한 동향을 살펴보고, 콘텐츠 적응에 속하는 중요 기술들 중에 비디오 트랜스코딩(video transcoding)과 모달리티 변환(modality conversion)에 대한 일반적인 사항과 우리의 연구 결과들을 전개하고자 한다. 비디오 트랜스코딩 관점에서는 최적의 트랜스코딩 연산 조합을 찾는 문제에 있어서 비트율-왜곡(rate-distortion) 모델(model)에 기초한 방법과 의미적 개념(semantic concept)이 판단에 미치는 영향에 대해 논한다. 모달리티 변환 관점에서는 최적의 모달리티 변환 경계를 찾기 위한 중첩 콘텐츠 값(overlapped content value, OCV) 모델을 논하고 실질적인 모델링 예제를 통해 OCV 모델의 효율성을 보인다.

I. 서 론

디지털 기술의 발달로 우리 일상생활은 방송, 유선통신, 무선통신 등 각 부문간에 인프라 및 서비스 차원의 융합에 의한 영향을 맞이하고 있다. 더욱이 이러한 환경의 전환과 함께 제공되는 방대한 디지털 멀티미디어 콘텐츠를 소비할 수 있는 PDA, Desktop PC, Laptop PC, PVR, Mobile phone 등의 다양한 단말기기가 보급이 되고, 이러한 단말기기 간의 통신이 가능한 광대역통합네트워크(Broadband convergence Network, BcN)와 같은 통합 네트워

크 환경이 조성되고 있다. 이러한 다변적 환경 속에서 방대한 멀티미디어 콘텐츠를 언제 어디서나 자유롭게 소비할 수 있는 기술적 요소들의 필요성이 대두되고 있다. 즉, 최근에 멀티미디어 소비 환경이 다양해짐으로써 사용자에게 멀티미디어 컨텐츠의 QoS(Quality of Service) 제공이 필수적인 요소로 부각되고 있다. 따라서 컨텐츠 제공자는 최종 사용자가 주어진 환경 자원에 가장 적합한 최대한의 품질을 가지는 멀티미디어 콘텐츠의 소비가 가능하도록 지원해야 할 필요가 있다.

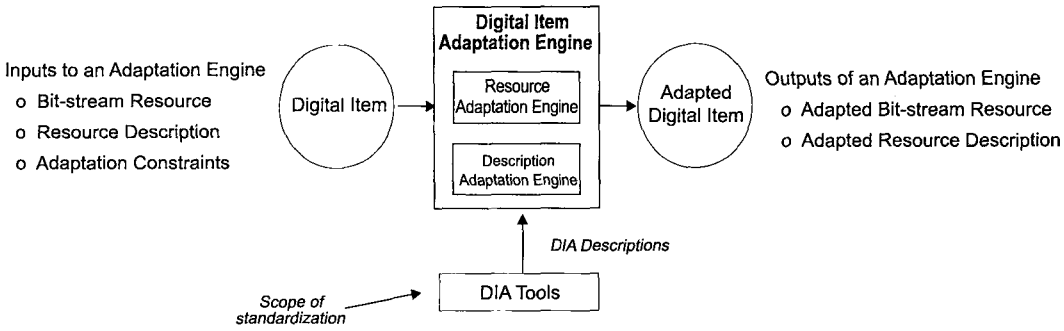
이를 위한 기술적 연구의 범주로써 우리가 매일 생활하는 환경과 디지털 기술을 적절하게 연결하기 위한 유비쿼터스 컴퓨팅(ubiquitous computing) 연구가 활발히 진행되고 있으며, 디지털 멀티미디어 콘텐츠를 다양한 환경에 적응(adaptation) 시킬 수 있는 기술이 필수적이다. 이를 위한 중요한 요소 기술로써 콘텐츠 적응 (content adaptation) 기술을 들 수 있다. 콘텐츠 적응은 크게 두 가지의 기술 범주로 구분할 수 있는데, 하나는 콘텐츠의 모달리티(비디오, 이미지, 오디오, 텍스트)의 변화 없이 주어진 제약조건(constraint)에 따라 콘텐츠 자체의 품질(quality)을 변화 시키는 콘텐츠 스케일링(content scaling)이며, 다른 하나는 콘텐츠가 가지는 모달리티를 다른 모달리티로 변환하는 모달리티 변환(modality conversion) 기술이다.

이러한 연구의 일환으로서 MPEG-21의 표준화 그룹에서는 디지털아이템 적응(digital item adaptation, DIA) 기술이 표준화 되었으며, 이는 다가올 융합 환경 하에서 범용적 멀티미디어 접근을(universal multimedia access, UMA) 가능하게 하는 시스템적인 방안을 제공할 것으로 기대된다 [1]. 이러한 맥락에서 본 논문은 콘텐츠 적응 기술에 있어서 MPEG-21의 표준화 동향을 살펴보고, 각 기술들에 대한 소개를 시작으로 콘텐츠 적응에 관한 연구들에 있어 비디오 트랜스코딩(video transcoding)과 모달리티 변환(modality conversion)에 대한 연구를 중점적으로 기술하고자 한다.

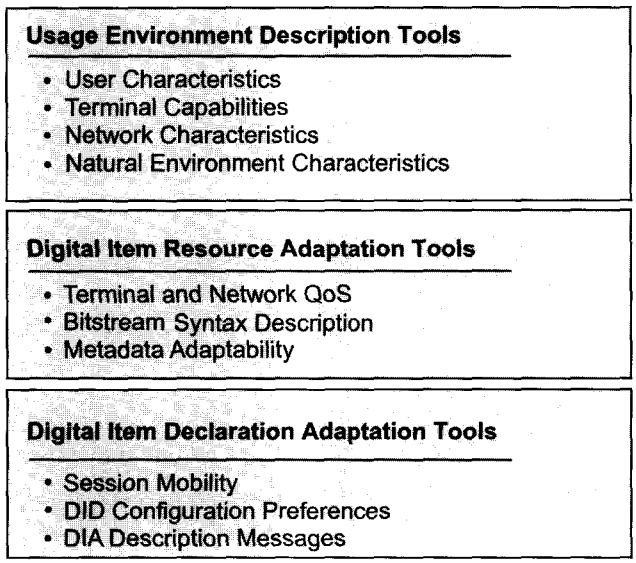
본 논문은 다음과 같이 구성된다. 2절에서는 MPEG-21 DIA 표준에서의 콘텐츠 적응에 대해 살펴본다. 3절에서는 콘텐츠 적응에 대한 기존 연구들과 구조적 측면에 대해 논하며, 4절과 5절에서는 콘텐츠 적응의 두 가지 중요한 측면인 비디오 트랜스코딩과 모달리티 변환 기술에 대한 우리의 연구 결과들을 기술하며, 마지막으로 6절에서 논문의 결론을 맺는다.

II. MPEG-21 콘텐츠 적응 (MPEG-21 DIA)

MPEG-21은 사용자가 구조화된 멀티미디어 콘텐츠인 디지털아이템(digital item)을 다양한 환경에서 상호 호환적으로 편리하게 생성, 교환, 소비할 수 있는 방법을 정의하고 실현할 수 있는 멀티미디어 프레임워크를 구축하는데 목적이 있다 [2]. 이 중 일곱 번째 요소인 디지털아이템 적응(digital item adaptation, DIA)은 디지털아이템을 사용자 특성 및 네트워크나 터미널의 특성을 고려한 적응을 가능하게 하고, 향상 시킴으로써 투명하고 증감된 접근을 제공하는데 목적이 있다 [1]. DIA는 그림 1과 같이 리소스(resource) 및 메타데이터를 적응하기 위한 서술(description) 체계와 포맷 비의존적인(format-independent) 적응 방법을 표준의 범위에 두고 있다. DIA는 그림 2와 같이 크게 세 가지 툴에 대한 표준을 정의 하고 있다. 첫째는 사용 환경에 대한 서술 체계이고, 둘째는 리소스 및 메타데이터를 사용 환경에 맞게 적응하기 위한 툴, 셋째는 세션 이동성과 디지털아이템 제공, 소비를 위한 설정 및 메시지 서술 체계를 정의하고 있다.

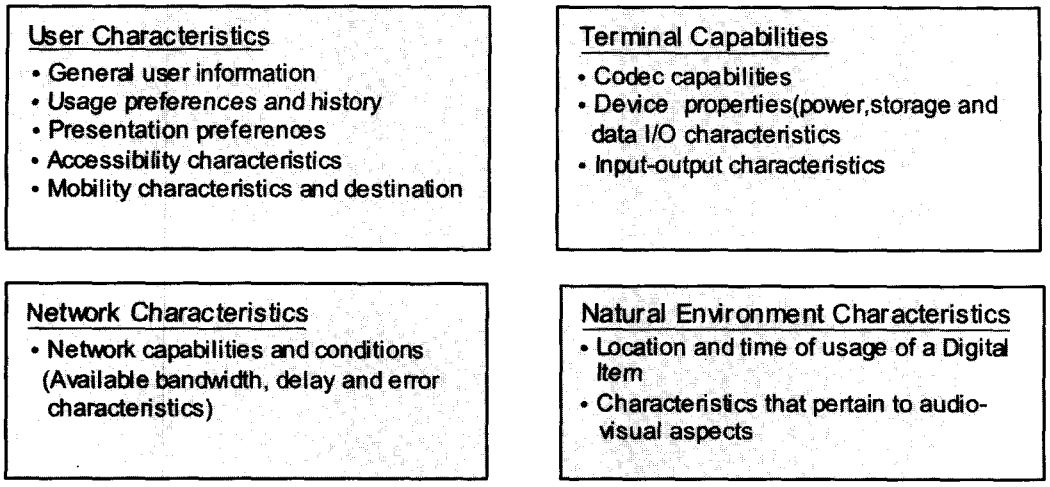


(그림 1) MPEG-21 DIA 개념 [1]



(그림 2) 디지털아이템 적응 툴

사용 환경 서술(usage environment description)에는 그림 3에서 보듯이 네 가지의 서술 체계가 존재한다. 첫째로, 사용자 선호도, 콘텐츠의 표현(representation) 방법에 따른 선호, 접근성(accessibility) 및 이동성(mobility)과 같은 다양한 사용자의 특성을 기술하기 위한 툴을 명명한 사용자 특성 (user characteristics) 서술자, 둘째로 터미널의 콘텐츠 부호화 특성, 전력 및 저장 용량 특성, 입출력 특성 등의 수행 능력 특성들을 기술한 터미널 능력 (terminal capabilities) 서술자, 셋째로 대역폭, 지연, 에러 특성과 같은 전송을 위한 네트워크의 수행 능력을 기술하는 네트워크 특성 (network characteristics) 서술자, 마지막으로 소비하고 있는 위치와 시간, 주위 환경 등을 기술하는 자연환경특성 (natural environment characteristics) 서술자로 구성된다.



(그림 3) 사용 환경 서술 툴 (Usage Environment Description Tools)

디지털아이템 자원 적응(digital item resource adaptation) 툴은 세 가지의 툴로 구성된다. 첫째로, 터미널 및 네트워크 QoS(terminal and network QoS)에서는 터미널들 및(또는) 네트워크의 특성에 기인한 제약조건(constraint)들을 만족시키기 위한 미디어 자원 적응에서 QoS (quality of service)를 최대화하는 최적의 적응 파라미터들을 선택하는 문제들을 언급하고 기술하고 있다. 즉, 이 툴은 (그림 4)에서 보듯이 환경 제약조건들을 만족하는 가용한 적응 연산들(adaptation operations)과 QoS 또는 유틸리티(utility)와의 상호 연관 관계를 명시할 수 있는 서술자들을 포함한다.

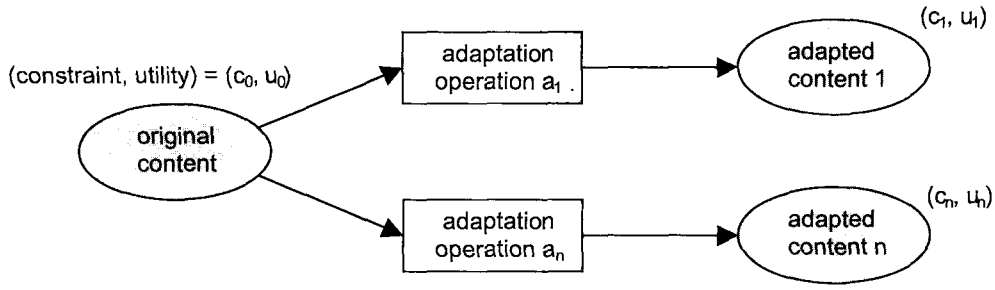
DIA 표준에서는 "AdaptationQoS" 서술자를 통해 상기한 상호 연관 관계를 기술 할 수 있으며, 이 서술자는 "UtilityFunction", "LookUpTable", 그리고 "StackFunction"으로 구성되어 있다. "Utility Function"은 리스트(list) 형식을 통해 제약조건들 및 유틸리티와 적응연산들간의 연관 관계를 표현할 수 있으며, "LookUpTable"은 매트릭스(matrix) 형태를 통해 다소 복잡한 연관 관계를 표현할 수 있고, "StackFunction"은 앞의 툴들에서의 나열식(enumerative)이 아닌 좀더 기능적인(functional) 관계 표현을 가능하게 한다.

둘째로, 비트스트림 신택스 서술(bitstream syntax description)에서는 부호화된 비트스트림 형태의 오디오 비주얼에 대한 상위레벨 구조를 XML 형식으로 서술함으로써 복호화 과정을 거치지 않고서도 XML 문서만을 변환(transform)하여 비트스트림을 쉽고 다이나믹하게 적응 가능하게 한다.

셋째로, 오디오/비주얼 콘텐츠 자체에 대한 적응뿐 아니라 디지털아이템 내의 메타데이타(metadata)의 적응과 관련된 정보들을 서술하도록 함으로써 적응과정에서 복잡도를 줄일 수 있도록 한다.

III. 콘텐츠 적응

본 절에서는 콘텐츠 적응에 대한 일반적인 사항과



(그림 4) 제약조건(c), 유틸리티(u), 적응연산(a) 간의 연관관계 서술 예 [1]

이에 대한 기존의 연구들을 살펴보고 2절에서 설명한 MPEG-21 DIA 프레임워크에 기반을 둔 콘텐츠 적응 시스템에 대해 논한다.

3.1 콘텐츠 적응 관련 기존 연구

대역폭과 단말기의 제한 조건에 관계없이 인터넷을 통해 최상의 정보 접근과 합의된 QoS를 제공하기 위해 UMA에 대한 연구가 최근 활발히 진행 중이다. 이러한 UMA 관련하여 콘텐츠 적응 (content adaptation) 연구의 세계 연구 현황은 다음과 같다.

IBM 연구팀은 "InfoPyramid"라고 불리는 다중 모달리티 계층과 다중 해상도 계층을 이용하여 PDA, 휴대용 컴퓨터, 이동전화, PC 등과 같은 퍼베이시브 디바이스(pervasive device)를 위한 UMA 서비스 제공에 초점을 맞추어 연구를 진행하였다 [3][4]. Hewlett-Packard의 "Cooltown" 프로젝트는 모바일(mobile) 환경에서 광범위한 e-service를 가능하게 하는 기술에 초점을 맞추고 있다 [5]. Columbia-ETRI 공동의 UMA 프로젝트에서는 콘텐츠의 유틸리티(utility)에 기반한 콘텐츠 적응에 대한 연구를 수행하였다 [6]. 또한 상호호환적인 UMA 시스템을 위해 멀티미디어기술구조(multimedia description scheme, MDS)에 대한 표준화를 시도한 MPEG-7과 함께 앞서 살펴본 것처럼 MPEG-21 표준화에서는 콘텐츠 적응을 자동화(automation)할 수 있는 프레임워크에 대한 연구가 진행되었다 [1][7]. 유럽의 "DANAE" 프로젝트 같은 많은 연구들이 국내외에서 MPEG 표준화와 더불어 생겨났으며 진행되고 있는 실정이다 [8][9].

3.2 콘텐츠 적응 시스템 구조

우리는 콘텐츠 적응 연구의 일환으로서 MPEG-21 DIA에 호환적인 클라이언트 서버 구조의 콘텐츠 적응 시스템을 구현하였다. (그림 5)는 사용자가 환경 특성에 따라 멀티미디어 콘텐츠를 자유롭게 소비할 수 있는 적응 변환 프레임워크를 위한 콘텐츠 적응 서버의 구조를 보여준다. 우리는 이러한 프레임워크를 바탕으로 "Active Room"이라 불리는 테스트베드를 구축하였다[10]. 그림에서 보듯이 적응 시스템은 (1) 제한조건검출 및 정보교환 엔진, (2) 적응변환 엔진, 그리고 (3) 저장소의 세 가지 하위 엔진으로 이루어지며, 각각의 역할을 간략히 살펴보면 다음과 같다. 제약조건 검출 및 정보교환 (constraints detection and exchange) 엔진은 네트워크 특성, 터미널 능력, 사용자 특성과 같은 환경 정보 및 세션 정보 (Session Information)를 수집함으로써 결정엔진

(decision engine)이 주어진 환경의 제약조건(constraint)에 따라 최적의 콘텐츠 적응 연산들을 선택할 수 있도록 정보를 제공한다.

적응 엔진 (adaptation engine) 모듈은 크게 판단엔진(decision engine)과 자원 적응 엔진(resource adaptation engine) 등과 같은 핵심 엔진들을 포함한다. 이 적응 엔진에서는 판단엔진의 결정에 따라 비디오 콘텐츠를 스케일링하는 비디오 트랜스 코딩과 비디오에서 이미지로의 변환과 같은 모달리티 변환 기능을 포함한다.

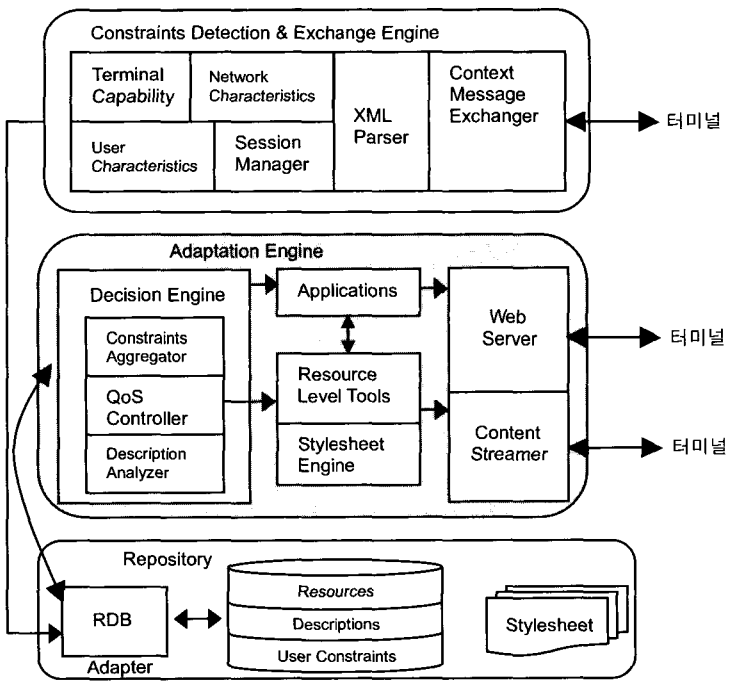
즉, 판단엔진은 제약조건 검출 및 정보교환 모듈로부터 제약조건에 관한 정보를 바탕으로 사용자에게 최상의 표현을 제공하기 위해 적절한 트랜스코딩 또는 콘텐츠 모달리티를 선택한다.

그리고 자원 적응변환 엔진은 판단엔진의 결정에 따라 실제적인 적응변환을 수행하는 엔진이다. 따라서, 비디오 트랜스 코딩 및 모달리티 변환을 위한 적응 알고리즘을 포함한다. 그리고 콘텐츠 적응 시스템에 사용되는 데이터베이스에는 기본적으로 멀티미디어 콘텐츠, 콘텐츠에 대한 서술(description), 그리고 사용자 정보 및 제약조건 정보들이 저장된다.

IV. 비디오 트랜스코딩 (Video Transcoding)

본 절에서는 3절에서 언급한 적응 엔진(adaptation engine) 중에서 비디오 트랜스코딩을 위한 판단 엔진(decision engine)에 대해 논한다.

사용자 환경에 맞는 최적의 품질 제공을 위한 최적



(그림 5) 콘텐츠 적응 서버 구조도

의 비디오 트랜스코딩 전략을 찾는 문제는 적응 변환된 개체의 품질을 최대화하기 위해 기본적으로 네트워크나 터미널, 그리고 사용자 특성에 의한 제한조건(constraint)들을 충족시켜 줄 수 있는 최적의 트랜스코딩 연산(operation) 조합(combination)을 결정하는 문제로 볼 수 있다.

예로, 입력된 영화 클립을 무선 환경의 모바일 폰을 소지한 사용자에게 서비스하기 위해 비디오 트랜스코딩 엔진에서는 무선 네트워크와 터미널 환경에 맞는 영화 클립을 재생산해서 전송할 필요가 있다. 이러한 트랜스코딩 판단을 위해 판단 엔진에서는 다음과 같은 질문의 답을 찾아야 한다.

"입력된 영화 클립은 주어진 사용자 환경 하에서 어떠한 트랜스코딩을 통해 최종 사용자에게 최상의 품질(quality)을 서비스할 수 있는가?"

비디오 비트율을 변환하는 트랜스코딩 연산에는 기본적으로 공간-시간-SNR(spatio-temporal-SNR)의 해상도(resolution)를 변화시키는 세 부류의 연산들이 있으며, 보다 효율적으로 자원(비트율)을 할당하기 위해서는 이러한 다차원적 트랜스코딩 연산들 상호간에 트레이드오프(tradeoff)를 고려할 필요가 있다.

예를 들어서, 전송 채널의 대역폭(bandwidth)이 감소하는 경우에 낮은 SNR(signal to noise ratio) 품질을 가진 컨텐츠를 좀 많이 전송하는 것과 높은 SNR 품질을 가진 컨텐츠를 좀 적게 전송하는 것 중에서 선택을 하거나, 때로는 작은 화면 크기의 프레임을 좀 많이 전송하는 것과 큰 화면 크기의 프레임을 좀 적게 전송하는 것 중에서 선택을 해야 한다.

위와 같은 비디오 트랜스코딩 판단 문제를 정형화 한다면 다음과 같이 비트율-왜곡(rate-distortion, R-D) 관계로 표현될 수 있다. R-D 최적화 문제는 트랜스코딩 된 콘텐츠의 평균 왜곡을 최소화하는 최적의 연산 조합 {RQ^*, TS^*, SS^*}를 찾는 문제이며, 다음 수식과 같이 정형화 할 수 있다.

$$\{RQ^*, TS^*, SS^*\} = \arg \min_{RQ,TS,SS} \sum_{i=1}^{L} D_i(RQ,TS,SS),$$
$$subject\ to \sum_{i=1}^{L} R_i(RQ,TS,SS) < R_{\max} \quad (1)$$

여기서 D_i는 i 번째 프레임의 왜곡(distortion)을 나타내고, R은 비트율을 나타낸다. 또한 $RQ=[RQ_1, RQ_2, \cdots RQ_L], RQ \in [RQ_{\min}, RQ_{\max}],\ i=1,...L,$, 여기서 RQ_i는 i 번째 프레임의 재양자화 (requantization) 파라미터이고, L은 시간적 세그먼트 (예, 비디오 장면)에서 전체 프레임(frame) 수를 나타낸다. 또한, $TS=[TS_1, TS_2, \cdots TS_L], TS_i \in \{False, True\}$, 여기서 TS는 시간적 스케일링(temporal resolution scaling) 연산 벡터이고, $False$ 는 비디오의 한 프레임이 버려짐(dropping)을 의미하며, $True$는 해당 프레임이 코딩됨을 의미한다. 그리고 $SS=[SS_1, SS_2, \cdots SS_L], SS_i \in [0,1]$, 여기서 SS는 공간적 스케일링(spatial resolution scaling) 연산 벡터이며, 예를 들어 SS=1은 원래의 프레임 크기를 갖는다는 의미하며, SS=0.25는 가로 그리고 세로의 프레임 크기가 각각 반으로 공간적 다운 스케일링되는 연산을 의미한다.

일반적으로 i 번째 프레임의 다차원 트랜스코딩에 의한 왜곡은 다음 식과 같이 각 차원의 연산에 의해 발생된 왜곡들의 가중치를 고려한 합으로 표현될 수 있다.

$$D_i(RQ,TS,SS) = w_{rq} D_i^{rq}(RQ,TS,SS) + w_{ss} D_i^{ss}(RQ,TS,SS) + w_{ts} D_i^{ts}(RQ,TS,SS) \quad (2)$$

여기서 D_i^{rq}, D_i^{rq}, 그리고 D_i^{ts}는 각각 재양자화, 시간적 스케일링, 그리고 공간적 스케일링 연산들에 의한 왜곡을 나타내며, $\{w_{rq}, w_{ts}, w_{ss}\} \in [0,1]$는 각각을 위한 가중치를 의미하며, 여기서 $w_{rq} + w_{ts} + w_{ss} = 1$.

이러한 트랜스코딩 판단 문제를 풀기 위해서 많은 연구들이 진행되어왔다. 기본적으로 비디오 트랜스코딩 판단 문제를 풀기 위해 품질(또는 왜곡) 측정이 중요한 문제로 부각된다. 즉, 트랜스코딩 판단을 위해서는 기본적으로 트랜스코딩 된 비디오의 품질 측정이 필요하며, 측정된 품질을 바탕으로 최상의 품질을 제공할 수 있는 트랜스코딩 연산들을 찾을 필요성이 있다. 품질 측정에는 크게 PSNR(peak signal to noise ratio)이나 MSE(mean square error)와 같은 객관적인 측정(objective measure) 방법과 인간의 주관적인 판단에 의한 측정(subjective measure) 방법이 있으며, 이러한 품질 측정을 바탕으로 비디오 트랜스코딩 판단 문제를 풀고자 시도한 많은 연구들이 있었다 [11][12][13][14]. 하지만 여기서 중요한 이슈로는 통계적 분석에 기초한 분석적 접근(analyti

cal approach)에 의한 품질 모델링[15]과 객관적 품질 측정과 주관적 품질 측정과의 차이를 극복하는 것과 여러 도메인의 (예로 공간 도메인, 시간 도메인, 의미적 도메인 등) 품질들을 하나의 품질로 취합하는 방법에 대한 연구가 필수적이다 [16][20].

콘텐츠 트랜스코딩의 과정은 비트율(일반적으로 자원(resource))에 따른 트랜스코딩 된 콘텐츠 품질의 관계를 나타내는 비트율-품질(rate-quality) 곡선에 의해 나타내 질 수 있다. UMA를 위한 최근의 연구 경향들은 앞서 논의했던 것처럼 콘텐츠 적응을 자동화(automation) 할 수 있는 이러한 비트율-품질 곡선을 이용하는 것이라 할 수 있다. 앞서 언급했듯이 MPEG-21 DIA는 이러한 적응 시스템의 모델화를 위해 AdaptationQoS같은 많은 서술 툴(description tool)들을 제공한다[2].

상기한 다차원(multidimensional) 비디오 트랜스코딩에 있어서 최적화 문제를 풀기 위해, [16]에서 우리는 비트율-왜곡(rate-distortion) 모델링에 의한 접근을 시도하였다. 또한 트랜스코딩 연산들 간의 상관(相關)성 분석을 통해, 상관 왜곡(dependency distortion)을 포함하는 개선된 왜곡 모델과 비트율 제어 알고리즘을 포함한다. 향상된 비트율-왜곡 모델을 사용해서 다차원(Spatio-SNR-Temporal) 해상도를 고려하는 비디오 트랜스코딩 시에 최적의 연산 조합들을 찾음으로써, 현재 주어진 제약 조건하에서 사용자에게 최적의 QoS를 제공할 수 있도록 한다.

비디오 트랜스코딩 판단 문제를 접근하는데 있어서 또 다른 하나의 연구 방향으로는 비디오의 분류(video classification)에 따른 트랜스코딩 전략을 적용하는 시도가 있었다[13][14]. 기존 연구들에서는 움직임(motion) 또는 공간적 정보(spatial detail)와 같은 저 레벨(low level) 특징(feature)에 따른 트랜스코딩 특성을 파악하여 같은 부류(class)의 비디오들에 같은 트랜스코딩을 적용함으로써 판단의 효율성을 높이고자 하였다. 하지만 비록 같은 시공간적 특성을 지닌 비디오라 할지라도 트랜스코딩에 있어서 사용자들이 다른 결과를 선호할 수도 있다는 점에서 기존 연구들에 단점이 존재한다. 예를 들어, 움직임과 공간적 정보가 많지 않은 특성을 가지는 대화(dialog) 비디오 클립과 이와 비슷한 저 레벨 특징을 갖는 교육 방송에서 주로 등장하는 텍스트가 오버레이(overlay) 된 비디오 클립에 대해서 사용자들은 각기 다른 트랜스코딩 선호도를 가진다. 즉, 비디오 콘텐츠가 어떤 의미(semantic)를 갖는 가에 따라 트랜스코딩 판단의 결과는 달라 질 수 있다. 사용자들은 의미적 개념(semantic concept)이 비슷한 비디오들에 대해 공통된 트랜스코딩 선호 특성을 가진다 [17]. 예로, 액션(action)을 많이 포함한 비디오와 대화(dialog) 장면이 주를 이루는 비디오에 대해 사용자들은 트랜스코딩 된 결과의 선호에 있어서 각각의 고유한 선호 특성을 가짐을 생각할 수 있다. 따라서 다른 의미적 개념을 갖는 비디오들에 각기 다른 트랜스코딩 전략을 적용함으로써 사용자에게 최적의 품질

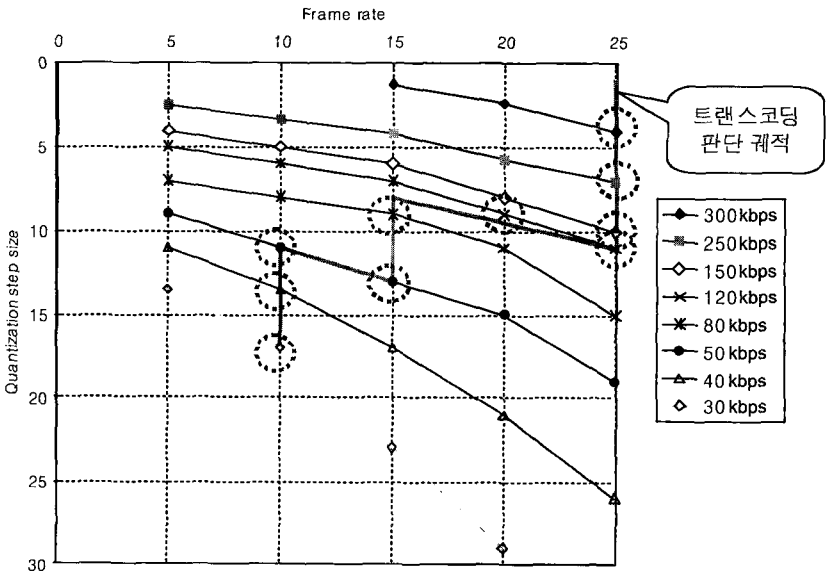
을 제공할 수가 있다.

이를 위해, 비디오들을 각각의 의미적 개념(semantic concept)에 따라 분류하고 각 종류별로 주관적인 테스트(subjective test)를 실시하여 트랜스코딩에 있어서 사용자의 시각적 인지(visual perception)의 특성을 분석한다. 주관적 측정을 통해 분석된 인간의 시각 특성들을 다차원 비디오 트랜스코딩 판단 시에 적용하기 위해서 트랜스코딩 판단 궤적(transcoding decision trajectory : TDT)을 모델화한다 [17]. (그림 6)에서와 같이 각 의미적 개념을 위한 궤적을 만들기 위해 우리는 각 특정 비트율에서 선택점(selection point)을 모두 표시하고 이 점들을 서로 연결함으로써 각 의미적 개념에 대한 TDT 선을 그린다. 트랜스코딩 판단 엔진에서는 이렇게 완성된 TDT를 이용함으로써 가장 좋은 트랜스코딩 연산의 조합을 찾아내는 결정을 내릴 수 있는 것이다. 어떤 의미적 개념을 갖는 비디오 부류들에 대해 주어진 많은 트랜스코딩 후보(candidate)들이 존재할 때 시스템은 궤적 선에서 가장 가까운 후보점을 검색함으로써 최적의 트랜스코딩 연산 조합을 간단히 결정할 수 있다[13][17].

V. 모달리티 변환 (Modality Conversion)

최근에 HD-TV, PDA, 핸드폰, MP3 플레이어등 다양한 멀티미디어 콘텐츠를 소비하는 단말기가 등장하고 있고 비디오, 오디오, 이미지, 그래픽 등의 다양한 콘텐츠 형식이 서비스되고 있다.

이런 다양한 콘텐츠와 다양한 디바이스를 가진 멀티미디어 소비환경에서는 기존의 단일 모달리티 내에서 만의 콘텐츠 스케일링(content scaling) 방법만으로 서비스질(QoS)을 조절 할 수 없다. 때로는 다른 모달리티로 컨텐츠를 변환하는 것이 필요할 수도



(그림 6) 트랜스코딩 판단 궤적 모델링

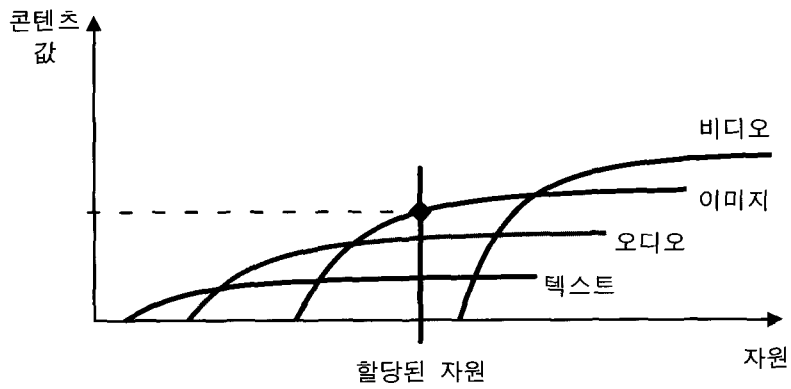
있다. 예를 들어, 대역폭이 너무 작을 때, 일련의 이미지 시퀀스를 보내는 것이 저화질의 비디오를 스트리밍하여 보내는 것보다 더 적합할 수 있다. 여기서 말하는 모달리티(modality)란 비디오, 이미지, 그래픽, 오디오, 텍스트와 같은 어떠한 양식(mode)뿐만 아니라, MPEG, JPEG, GIF와 같은 코딩 형식(format)들도 포함한다. 본 절은 QoS 측면에서 여러 모달리티를 고려하는 모달리티 변환에 관한 것이다.

모달리티 변환은 다양한 멀티미디어 소비 환경 하에서 콘텐츠의 모달리티 변환 방법을 제공함으로써 콘텐츠의 전달 및 소비에서 서비스 질을 유지할 수 있게 한다. QoS 관점에서 보면 모달리티 변환에서 중요한 현안은 "어떤 자원(resource) 특성 하에서 즉 제약조건(constraint) 하에서 현재의 모달리티가 어떠한 모달리티로 변환되어야 하는가?"이다. 즉 모달리티 사이의 변환 경계를 찾는 방법이 QoS를 위한 모달리티 변환에서의 핵심이라 할 수 있다. 따라서 본 절에서는 이와 같은 모달리티 사이의 변환 경계를 찾는 시스템적인 방법에 대해 논하고자 한다.

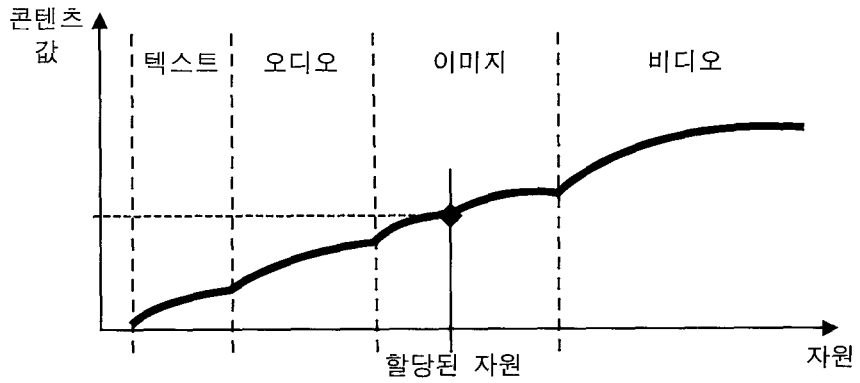
모달리티 변환의 판단에 영향을 끼치는 요소는 여러 가지가 있을 수 있으며, 우리는 크게 네 가지의 요소들로 분류하고자 한다. 첫 번째 요소로는 사용자의 소비환경에서 특정 모달리티를 지원하는가의 문제로 볼 수 있다. 예로, 사용자가 소지한 단말기가 텍스트만을 지원하는 무선기기일 경우 또는 주변 소음이 너무나 심한 경우는 오디오의 청취는 바람직하지 않을 수 있다. 두 번째 요소로는 사용자에게 있어서 특정 모달리티에 대한 선호도(preference)를 생각해 볼 수 있다. 세 번째 요소로는 자원 제약(resource constraint)이 될 수 있다. 예로, 비디오를 재생하기에는 충분한 네트워크 연결을 보장하지 않는 경우에도 모달리티 변환에 영향을 끼칠 수 있다. 네 번째 요소로는 콘텐츠가 담고 있는 의미 정보일 수 있는데, 예로서 인터뷰(interview) 비디오와 발레(ballet) 비디오인 경우에 전자는 텍스트(text)로의 변환이 적합할 수 있으나 후자는 그렇지 않을 수도 있다.

5.1 모달리티 변환 모델링 (Modeling Modality Conversion)

모달리티 변환을 위해서는 기본적으로 각 모달리티마다 서로 다른 성격을 갖는 품질들을 비교할 필요



(그림 7) 중첩 콘텐츠 값 (overlapped content value) 모델



(그림 8) 최종 콘텐츠 값 함수 (content value function)

성이 있다. 즉, 각 모달리티마다 생성된 비트율-품질(rate-quality) 곡선들을 한곳에서 비교함으로써 변환 경계점을 찾을 수 있다. 이러한 맥락에서 우리의 지난 연구에서는 체계적인 모달리티 변환을 위한 콘텐츠 값(content value 또는 quality)과 자원(resource) 및 모달리티 사이의 관계를 명확히 나타내주는 중첩 콘텐츠 값(overlapped content value, OCV) 모델을 제안하였다[18].

(그림 7)은 하나의 비디오 콘텐츠에 대해 여러 모달리티의 비트율-품질(rate-quality) 곡선(modality curve)들로 구성된 OCV 모델을 나타낸다. 모달리티 곡선(modality curve)의 교차점(intersection point)은 (그림 8)에서의 점선처럼 모달리티 사이의 변환 경계(conversion boundary)를 나타낸다. 이런 변환 경계에 기반 해서, 허용될 수 있는 자원의 범위 내에서 최대한의 QoS를 유지하는 모달리티 변환을 정량적으로 결정할 수 있다. 모달리티 j의 콘텐츠 값 곡선을 VM_j 라 하자. 여기서 j=1...J 이며, J 는 콘텐츠가 취할 수 있는 모달리티의 수를 의미하며, R은 자원(resource)을 의미한다. 또한 모든 j=1...J 에 대해 $VMj \geq 0$을 만족한다. 이때 콘텐츠 값 함수(content value function)는 다음 수학식과 같이 표현된다.

$$V = \max\left\{VM_j(R) \mid j = 1...J\right\}. \quad (3)$$

하나의 모달리티를 위한 콘텐츠 값 곡선(즉, 모달리티 곡선)은 여러 품질(quality) 측정 방법에 의해 측정될 수 있다.

예를 들어, 비디오 모달리티는 PSNR (peak signal to noise ratio) 또는 MOS (mean opinion score)의 계산에 의해 콘텐츠 값 곡선을 각각 얻을 수 있다. 이러한 방식에 의해 적절히 구해진 각 모달리티 곡선들을 (그림 8)에서처럼 OCV 모델로 표현하여 QoS에 따른 모달리티 변환점을 결정할 수 있다.

우리는 [18]에서 이러한 OCV모델은 MPEG-21 DIA의 변환 선호도(conversion preference) 툴을 이용하여 기술된 특정 모달리티 변환에 대한 사용자의 선호도를 반영할 수 있는 바탕이 될 수 있음을 보였다[1].

5.2 모달리티 변환 예제 (Example of Modality Conversion)

이번 절에서 우리는 하나의 터미널에서 네트워크를 통해 스트리밍되는 하나의 비디오에 대한 모달리티 변환에 대한 가능성을 살펴본다. 실험에서는 119.3Kbps, QCIF, 25fps의 해상도를 갖는 300 프레임들로 구성된 Foreman MPEG-4 테스트 비디오를 사용하였다.

실제적인 모달리티 변환은 오프라인(offline)에서 구동되며, 특정 비트율 제약조건 하에서, 콘텐츠 적응 시스템은 이에 적절한 품질과 모달리티를 갖는 버전을 선택한다. 적응된 비디오 버전들을 생성하기 위해 프레임 버려짐(frame-dropping)과 재양자화(requantization) 연산의 조합을 사용하고, 이미지 시퀀스(image sequence)들은 [19]에서의 방법을 이용해 키프레임들로서 추출된다. 추출된 이미지들은 MPEG 비디오의 I 프레임과 같이 JPEG 포맷으로 인코딩된다.

그리고 원본 콘텐츠의 서술(description)에 해당하는 텍스트(text) 버전은 인위적으로 작성되었다. 오디오 버전은 기술된 텍스트로부터 음성(speech)로 생성되었으며, XingMPEG® 인코더를 사용해 24kbps and 8kbps를 갖는 두 가지 버전으로 생성되었다. Foreman 비디오를 위한 모든 적응 변환된 버전들은 〈표 1〉과 같다.

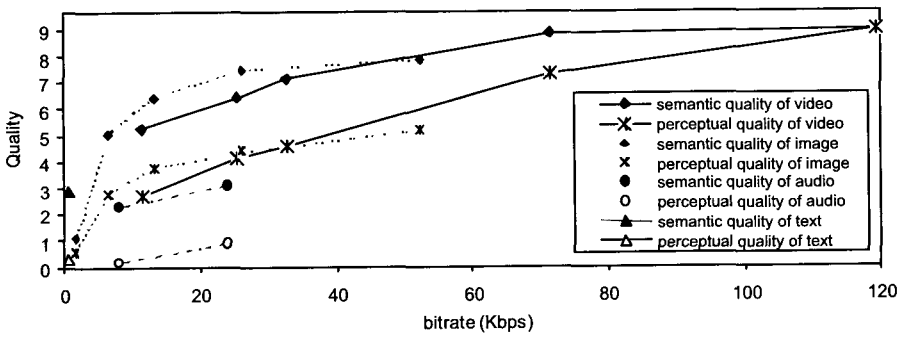
〈표 1〉 Foreman 비디오의 적응된 콘텐츠 버전들. 여기서 f는 프레임율(frame rate)를 의미하며, q는 양자화(quantization) 파라미터를 의미한다.

순서	모달리티	비트율 (Kbps)	설명(description)
1	Video	119.33	Original video, q = 10, f = 25fps
2	Video	71.33	Dropping all B frames, q = 10, f = 8.3fps
3	Video	32.67	Dropping all B and P frames, q =10, f = 1.7fps
4	Video	26.33	Dropping all B frames, q = 30, f = 8.3fps
5	Video	11.33	Dropping all B and P frames, q = 30, f =1.7fps
6	Image	52.27	Sequence of 32 images, q = 10
7	Image	26.14	Sequence of 16 images, q = 10
8	Image	13.07	Sequence of 8 images, q = 10
9	Image	6.53	Sequence of 4 images, q = 10
10	Image	1.63	Sequence of 1 image, q = 10
11	Text	0.5	A stream of explanatory text
12	Audio	24	Explanatory speech with high quality
13	Audio	8	Explanatory speech with low quality

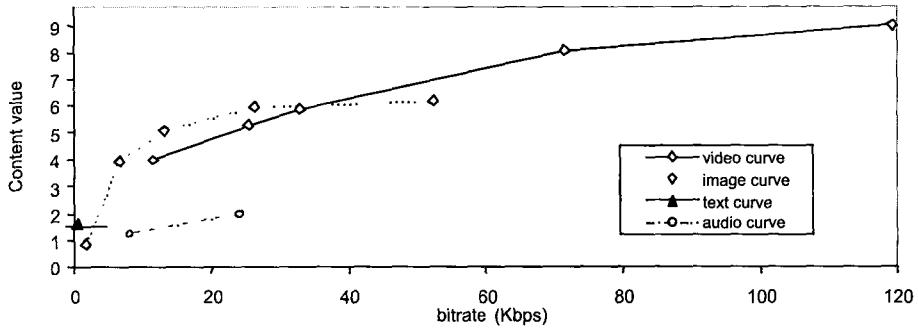
우리는 콘텐츠 값(content value)을 측정하기 위해 인지적 품질(perceptual quality)과 의미적 품질(semantic quality)로 나누어서 생각한다.

전자는 사용자의 시각적/청각적인 인지도에서의 만족 정도를 나타내며, 후자는 전달되는 의미적 정보(information)의 양과 관계가 있다.

그리고 최종 콘텐츠 값은 두 가지 품질의 평균으로서 정의된다. 인지적 품질과 의미적 품질은 사용자의



(a)



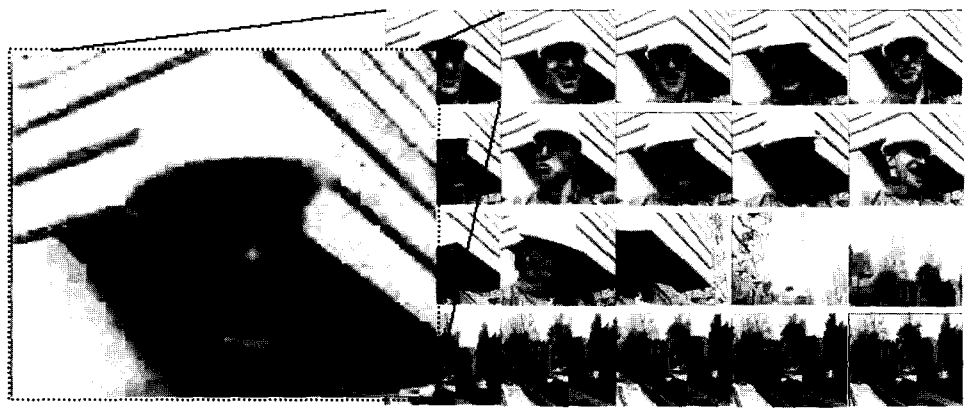
(b)

(그림 9) Foreman 비디오를 위한 (a) 모달리티들의 품질곡선(quality curve)과 (b) OCV 모델

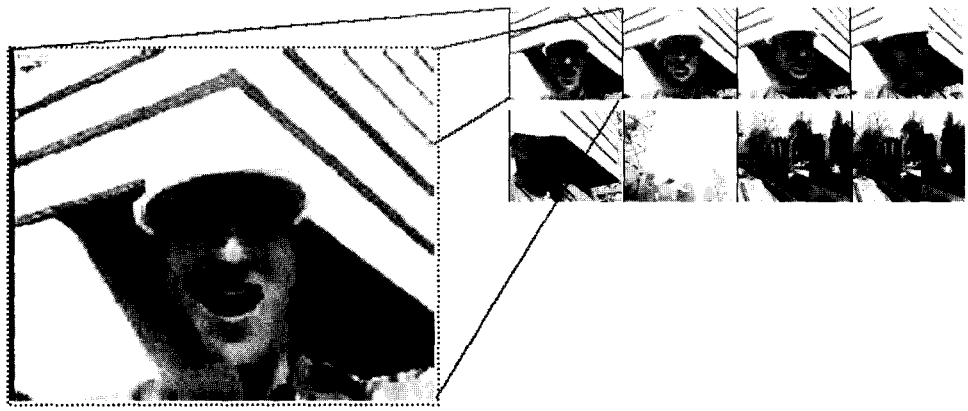
주관적인 테스트를 통해 얻어질 수 있는 데이터에 대한 자세한 절차 및 결과 분석은 [20]에서 논하였다. (그림 9a)는 여러 모달리티들에서의 인지적 품질(perceptual quality)과 의미적 품질(semantic quality) 곡선들을 보인다.

이렇게 얻어진 두 가지 곡선들의 평균을 취함으로써 최종적으로 구해진 OCV 모델을 (그림 9b)에 나타내었다. 그림의 OCV 모델을 확인함으로써 32kbps에서 비디오와 이미지 모달리티간의 변환점과 2.4kbps에서 이미지와 텍스트간의 변환점이 존재함을 알 수 있다.

(그림 10)은 13.5kbps에서 비디오와 이미지의 두 가지 콘텐츠 버전들을 보이고 있다. 비디오 버전은 30 양자화값으로 인코딩된 20장의 I 프레임들로 구성되어 있으며, 이에 비해 이미지 버전은 원래의 공간적 품질을 가지면서 8장의 이미지들로 구성된다.



(a) 20장의 프레임들로 구성된 비디오 스트림



(b) 8장의 이미지들로 구성된 이미지 시퀀스(sequence)

(그림 10) 13.5kbps 비트율에서 비디오와 이미지 모달리티를 갖는 콘텐츠의 시각적 비교

(그림 11) 2kbps 비트율에서 이미지와 텍스트 버전들의 비교

그림에서 보듯이 적은 수의 중요 프레임들로 구성된 이미지 버전이 비디오 버전보다 좋은 품질로 나타남을 확인할 수 있다. 따라서 이러한 비트율 제약조건 하에서는 사용자에게 이미지 버전을 선택해서 전송하는 것이 합당함을 알 수 있다.

또한 비트율이 현격히 낮아진 경우(예로 2kbps)에 텍스트 버전이 선택됨을 (그림 9b)를 통해 알 수 있었다. (그림 11)은 이 경우에 각 콘텐츠들을 시각적으로 비교한 것이다. 이때의 이미지 버전은 단지 한 장으로 구성되어 있으며, 텍스트의 경우는 한 장의 이미지보다 더 많은 정보를 가질 수 있음을 확인할 수 있다.

위의 실험에서 보듯이, 제안한 OCV 모델을 사용함으로써(그림 9b) 우리는 Foreman 비디오를 비트율에 따라 적절한 모달리티로 변환할 수 있음을 살펴보았다.

VI. 결론 및 향후 과제

본 논문에서는 방송통신 융합 환경하에서 필수적인 UMA 서비스를 위한 콘텐츠 적응 기술에 대한 표

준화 동향과 기술적 검토를 논하였다. 표준화 동향에서는 MPEG-21 DIA에 대한 전체적인 소개를 하였으며, 콘텐츠 적응 관련 기술들에서는 비디오 트랜스코딩과 모달리티 변환에 대한 연구들을 언급하였다.

향후에는 방송 통신 융합 환경에서 사용자 특성(Quality of Experience)을 고려한 미디어 관점의 QoS와 네트워크 관점에서의 QoS를 유기적으로 통합하는 연구가 필요할 것이다. 또한 진화하고 융합되는 멀티미디어 통신환경에서 콘텐츠 적응 기술의 중요성 및 유용성을 검증하기 위하여 본 연구들을 방송통신 융합 환경에서 IP 기반 TV같은 구체적인 응용 기술로 적용이 필요하리라 생각된다.

[참고문헌]

[1] ISO/IEC 21000-7 FDIS Part 7: Digital Item Adaptation, ISO/IEC JTC1/SC29/WG11/N6168, Hawaii, USA, Dec. 2003

[2] ISO/IEC 21000-7 Part 1: Multimedia Framework, ISO/IEC JTC1/SC29/WG11/N4333, Sydney, July 2001

[3] R. Mohan, J. R. Smith, C.-S. Li, "Adapting Multimedia Internet Content for Universal Access", IEEE Trans. Multimedia, Vol. 1, No. 1, pp. 104-114, Mar. 1999

[4] J. R. Smith, R. Mohan and C.Li, "Scalable Multimedia Delivery for Pervasive Computing", ACM Multimedia' 99, 1999

[5] http://www.coooltown.hp.com/cooltown

[6] J.-G. Kim, Y. Wang, and S.-F. Chang, "Content-adaptive Utility-based Video Adaptation", In Proc. ICME-2003, Baltimore, July 2003

[7] ISO/IEC 15938-5 FDIS Part 5: Multimedia Description Schemes, ISO/IEC JTC1/SC29/WG11/N4242, Oct. 2001

[8] DANAE Website: http://danae.rd.francetelecom.com

[9] A. Perkis, J. Zhang, "Multimedia resource adaptation for streaming media", ACM, 2000

[10] Y. J. Jung, T. C. Thang, J. Lee, and Y. M. Ro, "Visual Media Adaptation System for Active Media," Int. Conf. Imaging Science, Systems, and Technology (CISST' 03), Vol. 2, pp. 405-409, June 2003

[11] P. A. A. Assuncao, and M. Ghanbari, "Optimal transcoding of compressed video," in: Proceedings of the ICIP 1997, October 1997, pp. 739-742

[12] T. Yoshida, T. Warashina, and Y. Inazumi, "Video transcoding based on optimal frame rate estimation," in: Proceedings of the ICIP 2003, September 2003, pp. 181-184

[13] N. Cranley, L. Murphy, P. Perry, "User-Perceived Quality-Aware Adaptive Delivery of MPEG-4 Content," in Proc. NOSSDAV' 03, pp. 42-49, June, 2003

[14] Y. Wang, S.-F. Chang, A. C. Loui, "Subjective Preference of Spatio-Temporal Rate in Video Adaptation Using Multi-Dimensional Scalable Coding," in Proc. ICME, June, 2004

[15] Z. He, S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," IEEE Trans. Circuits Syst. Video Technol. 11 (December 2001) 1221-1236

[16] Y. J. Jung and Y. M. Ro, "Joint control for

hybrid transcoding using multidimensional rate distortion modeling," in Proc. ICIP2004, Oct. 2004

[17] Y. J. Jung, Young Suk Kim, Duck Yeon Kim, Jea-Gon Kim and Young Man Ro, " Analysis of Human Perception for Semantic Concept-based Video Transcoding" IWAIT 2005, Jeju, Korea, 2005

[18] T. C. Thang et al., "CE Report on Modality Conversion Preference Part-I," ISO/IEC JTC1/SC29/WG11 M9495, Pattaya, Mar. 2003

[19] H.-C. Lee, S.-D. Kim, "Iterative Key Frame Selection in the Rate-Constraint Environment," Image Communication, Issue 18, pp. 1-15, 2003

[20] T. C. Thang, Y. J. Jung, Y. M. Ro, "Modality conversion for QoS management in Universal Multimedia Access," IEE Proceedings - Vision, Image and Signal Processing. (In press)

노용만

1985년 연세대학교 전자공학과 학사
1987년 한국과학기술원 전자 공학부 석사
1992년 한국과학기술원 전자 공학부 박사
1987년 컬럼비아대학 연구원
1992년 UC 어바인대학 초빙 연구원
1996년 UC 버클리대학 연구원
1997년 ~ 현재 한국정보통신대학교 공학부 부교수
관심분야 : 이미지/비디오 처리 및 분석, MPEG-7/21, 특징 인식, 이미지/비디오 인덱싱, Watermarking

정용주

1999년 홍익대학교 컴퓨터공학과 학사
2000년 한국정보통신대학교 공학부 석사
2000년 ~ 2001년 ㈜아이브이시스템 연구원
2000년 ~ 현재 한국정보통신대학교 공학부 박사 과정
관심분야 : 이미지/비디오 처리 및 분석, MPEG-7/21, 콘텐츠 적응, 비디오 코딩 및 영상통신

Truong Cong Thang

1997년 Hanoi University of Technology, Hanoi, Vietnam
2000년 MS in Hanoi University of Technology, Hanoi, Vietnam
1997년 ~ 1999년 Satellite engineer, Vietnam Telecom International
2001년 ~ 현재 Ph.D student in Information and Communications Univ.
관심분야 : Image/Video processing, Video abstraction, Content adaptation, MPEG-21