

지각도 모델에 근거한 UMA 를 위한 영상 변환 기법

이건섭^o 김유남 설상훈
 고려대학교 전자공학과
 {leeks,kyn,sull}@mpeg.korea.ac.kr

Perceptual Model-Based Image Transcoding for UMA

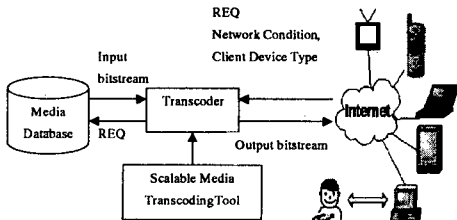
Keansub Lee^o Yunam Kim Sanghoon Sull
 School of Electrical Engineering Korea University

요 약

본 논문에서는 인간의 시각적인 감각을 멀티미디어 콘텐츠의 UMA 변환 서비스에 적용하여 영상의 다양한 디스플레이 크기의 사용자 단말기에 맞게 효율적으로 변환(해상도 축소나 Cropping) 기법을 제안하고 있다. 즉, 영상의 중요한 객체를 사각형 경계박스로 표시한 후 각각의 객체의 저자의 의도대로 사용자가 지각적으로 인식할 수 있는 최소의 공간 해상도 축소 정보를 정의하여, 영상의 변환 시 각각 객체를 사용자가 충분히 인식할 수 있는 한계치로 사용하여 효율적인 UMA 서비스를 보장하는 사용자 자원 재분배 기법을 제안한다. 또한, 본 논문에서 제안된 알고리즘을 기존의 방식과 비교하여 실험적으로 그 장단점을 비교한다.

1. 서론

최근에, 다양한 사용자 단말기와 네트워크의 발달로 멀티미디어 콘텐츠를 접근/이용하고자 하는 사용자 자원은 다양해지고 있고, 그러한 사용자 자원(client resource)에 맞게 원래의(original) 멀티미디어 콘텐츠를 다양한 매체 형태와 크기로 변환(transcoding)하는 서비스가 요구되고 있다. Universal Multimedia Access (UMA) [1]는 다양한 네트워크 상태, 사용자와 출판자의 선호도(preferences)나 단말기의 능력(capabilities)에 맞게 멀티미디어 콘텐츠(image, video, audio 등)를 변환, 전송함을 목적으로 한다(그림 1). 멀티미디어 콘텐츠의 효율적인 UMA를 수행하기 위해, 스케일링(scaling), 예를 들면 형태(format)나 해상도(resolution)의 변화, 데이터 크기(in bits) 축소, 객체 제거(dropping) 등과 같은 방법(mechanism)이 사용된다[2].



< 그림 1. 멀티미디어 콘텐츠 UMA 서비스 >

본 논문은 영상내의 각각의 객체(object)나 영역(region)의 공간적인 해상도 축소 정보, 즉 인간의 지각도(perceptibility)를 만족하는 최소한의 공간적인 해상도를 제안한다. 즉, 그 영상을 처음에 저자나 출판자가 의도했던 목적대로 사용자가 정확히 인지할 수 있는 한계점(threshold)을 표현하는 값인 공간적인 해상도 축소 정보(spatial resolution reduction hint)를 정의 및 적용함으로써, 영상을 다양한 디스플레이 크기의 사용자 단말기에 맞게 효율적으로 변환(해상도 축소나 Cropping) 서비스를 할 수 있다.

그리고, 본 논문의 구성은 다음과 같다. 우선 2 장에서는 기존에 연구되고 있는 UMA 관련 연구 사례를 살펴보고, 3 장에서는 본 논문에서 제안하는 영상 변환 기법인 PMIT (Perceptual Model-Based Image Transcoding)의 개념과 적용 알고리즘을 소개한다. 4 장에서는 그 실험결과와 고찰에 대해서 알아보고, 마지막으로 최종적인 결론과 향후 연구과제를 5 장에서 기술한다.

2. 관련연구

John R. Smith 와 Rakesh Mohan 이 제안한 보편적인 접근(universal access)을 위한 멀티미디어 콘텐츠의 선택 방식[3,4,5]에서는 멀티미디어 콘텐츠를 멀티모달과 멀티해상도로 나타낼 수 있는 계층적 구조의 정보피라미드식 표현기법과 사용자 단말기 환경에 최대도 적합한 콘텐츠 표현 방식을 선택할 수 있는 기법을 제안하고 있다.

* 본 연구는 한국과학재단 특정기초연구(98-0102-04-01-3) 지원으로 수행되었음.

또한, 영상을 종류(type)와 목적(purpose)에 따라 분류한 후 그것에 따라 효율적으로 영상을 변환/전송한다. 예를 들면, 글자(text), 영상, 동영상, 음성의 매체 형식들과 그것들의 다양한 해상도/크기의 표현 방식들을 정의하고 사용자 단말기의 환경에 따라 선택적으로 멀티미디어 매체형식과 해상도를 택해서 사용자에게 변환/전송케 하는 방식이다. 원래의 영상의 크기가 256*256 인 경우에 사용자 단말기 디스플레이 사이즈가 적어짐에 따라 영상의 해상도를 축소하거나 글자로 변환해서 전송한다. 그리고, 영상 내의 다양한 의미 있는(semantic) 객체에 중요도(importance)를 부여하여 선별적으로 변환/전송할 수 있도록 한다. 중요도란 영상내의 객체들의 서로간의 상대적인 중요성을 나타내는 것으로, 0 ~ 1 의 값을 갖는다. 여기서, 1 은 가장 중요한 객체를 의미하고, 0 은 가장 덜 중요한 객체를 나타낸다[6, 7, 8].

그러나, 위의 제안된 방법은 허용할 수 있는 변환의 정도를 지각도에 근거한 수치적인 방법(measure)으로 표현하지 못했다. 비록 중요도가 영상의 어느 부분을 선별적으로 전송할 것인지에 대해 사용될 수 있지만, 각각의 중요 영역의 해상도를 얼마나 줄일 수 있으며 사용자가 인식할 수 있느냐에 대한 인간의 지각적인 정보를 제공하지는 못한다.

3. 공간적 해상도 축소 정보(ResReduction Information)

3.1 정의

만약에 사용자 단말기의 디스플레이 크기가 영상의 크기보다 작을 경우, 영상은 subsample 되거나 crop 해서 단말기의 공간적 해상도에 맞게 변환해야 한다. 이 경우에 과도한 해상도의 축소 때문에 사용자는 글자나 얼굴을 인식하지 못할 수도 있다. 그러므로, 변환된 영상을 처음의 저자나 출판자가 표현하고자 했던 의도대로 사용자가 이해할 수 있거나, 저자나 출판자가 보존되고자 했던 영상의 각각의 객체나 영역의 지각도의 정도(degree)를 보장해주는 정보는 매우 유용하다. 여기서, 객체는 영상내의 사각형(rectangular)의 박스(bounding box)에 의해 제한되는 공간적인 영역을 의미하며, 자동차나 다리, 얼굴, 글자 등의 의미 있는 영역으로 정의한다.

공간적 해상도 축소 정보는 영상내의 각각의 객체나 영역의 최대의 공간 해상도 축소 비율을 나타낸다. 그것은 저자나 출판자에 의해서 결정되며, 저자의 의도대로 사용자가 지각적으로 인식할 수 있는 최소의 공간 해상도 정보를 나타낸다. 만약에 영상내의 전체의 객체의 개수를 N 라고 하면, 그 영상의 공간적 해상도 축소 정보 R 은

$$r = \{ r_1, r_2, \dots, r_N \}$$

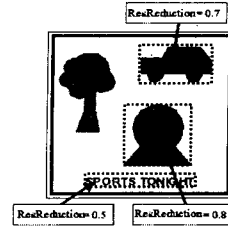
으로 표현되며, n 번째 객체의 공간 해상도 축소 정보 r_n 는 다음과 같이 정의될 수 있다.

$$r_n = r_n^{\min} / r_n^o$$

여기서 r_n^{\min} 는 n 번째 객체를 저자의 의도대로 사용자가 지각적으로 인식할 수 있는 최소의 공간 해상도이며, r_n^o 는 그 객체의 원래의 해상도로 정의한다. 그리고 변환 영상의 가로(width)/세로(height)비를 원래 영상의 가로/세로 비와 같은 경우는

$$r_n = W_n^{\min} / W_n^o = H_n^{\min} / H_n^o$$

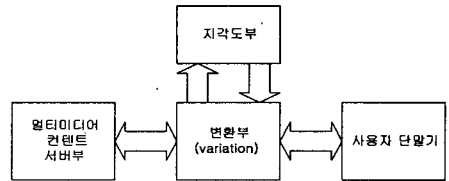
로 정의될 수 있다. 그리고 r_n 의 값은 0 ~ 1 이며, 여기서 1 은 객체의 원래의 해상도를 의미한다. 예를 들어, 100 * 100 의 객체의 R 값이 0.7 인 경우라면, 그 객체는 사용자의 지각도를 떨어뜨리지 않고 70 * 70 의 해상도의 크기까지 줄일 수 있다.



< 그림 2. 공간적 해상도 축소 정보의 예 >

3.2 사용자 자원 재분배(resource allocation) 알고리즘

우선, 지각도를 이용한 멀티미디어 변환 방법 및 장치에 관한 전체적인 시스템 개략도를 소개한다. 멀티미디어 콘텐츠 서버부(multimedia content server)에서는 원래의 멀티미디어 콘텐츠 뿐만 아니라, 미리 변환된 다양한 멀티미디어 콘텐츠를 저장/관리한다. 사용자 단말기부(client device)에서는 사용자 단말기의 자원을 관리하며 멀티미디어 콘텐츠 서버에 콘텐츠를 요구 시에 그 단말기 자원 정보를 같이 전송한다. 변환부에서는 사용자 단말기에서 요구가 있을 때, 멀티미디어 콘텐츠 서버에서 적합한 멀티미디어 콘텐츠를 찾아서 전송하거나 사용자 단말기 자원에 맞는 다양한 매체형태 및 프레임 해상도와 비율로 변환해서 사용자 단말기에 전송한다. 이때 지각도부에 있는 그 멀티미디어 콘텐츠에 해당하는 지각도 정보를 이용한다. 즉, 지각도부에서는 변환부에서 멀티미디어 콘텐츠를 변환시에 그 콘텐츠에 해당하는 지각도에 관한 정보를 제공한다.



< 그림 3. 지각도를 이용한 영상 변환 서비스의 개략도 >

우리는 영상의 변환을 다음과 같은 자원 재분배 알고리즘으로 모델링 할 수 있다. 만약, 영상의 i 번째 객체의 중요도와 해상도를 각각 S_i, r_i 라고 하면,

$$\text{Max}(V(I, r)) = \text{Max} \left(\sum V_i(r) \right) = \text{Max} \left(\sum \min(1, r) \cdot S_i \cdot u(r - r_i) \right)$$

$$\text{such that } \begin{cases} r | x_f - x_i | \leq \text{width}_{\text{client}} \\ r | y_f - y_i | \leq \text{height}_{\text{client}} \end{cases}$$

$$u(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

여기서, (x_i, y_i) 와 (x_f, y_f) 은 각각 사용자 단말기의 디스플레이 될 영상의 시작점과 끝점으로 나타내는 사각형 영역을 의미하고, 이 영역에 포함된 객체들의 집합을 I, r 은 그 영역의 공간 해상도를 나타낸다. 즉,

$$r = \max \left(\frac{W}{|x_f - x_i|}, \frac{H}{|y_f - y_i|} \right)$$

전체 영상 변환 구조 C 는,
 $C = \{I, r\}$.

$V(I, r)$ 함수는 다음과 같은 성질을 가지고 있다.

$$\begin{aligned} &\text{고정된 } I, V(I, r_1) \leq V(I, r_2) \text{ if } r_1 < r_2, \\ &\text{고정된 } r, V(I_1, r) \leq V(I_2, r) \text{ if } I_1 < I_2. \end{aligned}$$

또한, 주어진 I 에서 최대의 r 은,

$$r_{\max}^I = \min_{i,j \in I} r_{ij} = \min_{i,j \in I} \left(\max \left(\frac{W}{|x_j - x_i|}, \frac{H}{|y_j - y_i|} \right) \right)$$

이고, V 는 $C(I, r_{\max}^I)$ 에서 최대값을 갖는다.

그러므로, 모든 가능한 $I \subset \{1, 2, \dots, N\}$ 에서, r_{\max}^I 와 $V(I, r_{\max}^I)$ 를 계산하여 적절한 C_{opt} 를 구한다.

4. 실험 및 고찰

4.1 실험환경과 결과

Test 영상

Source : MPEG-7 Content (News.mpg)

공간적 해상도 : 352 * 288

Red box : important and ResReduction 정보를 나타냄.

Workstation : 256 x 208	Color PC : 192 x 156	TV browser : 128 x 104	HHC : 96 x 76	PD A: 64 x52
				 (a)
				 (b)
				 (c)
				 (d)

<그림 4. (a)(c) ResReduction 정보를 사용하지 않은 경우
(b)(d) ResReduction 정보를 사용한 경우 >

4.2 고찰

Test 영상 (a)(b)의 경우는, 얼굴과 글자에 해당하는 중요한 2개의 객체가 각각 $R=0.47, I=1.0$ 과 $R=0.44, I=0.9$ 의 값을 갖는다. (a)의 경우에는 가장 중요한 객체인 얼굴이 HHC의 디스플레이 크기부터는 식별할 수가 없고, 글자의 경우에는 TV

browser 정도부터 식별할 수가 없다. 그러나, (b)는 각각의 객체를 식별할 수 있는 한도인 R 값을 사용함으로써 디스플레이 되는 모든 객체, 특히 얼굴을 정확히 식별할 수 있다.

Test (c)(d)는 3개의 얼굴인 객체로 영상의 왼쪽으로부터 각각 $I=1.0, R=0.54, I=0.8, R=0.6, I=0.7, R=0.45$ 의 값을 갖고, (c)의 경우에는 color PC의 정도부터 얼굴 식별이 불가능하지만, (d)의 경우는 보다 덜 중요한 객체부터 제거해 나감으로써 보다 중요한 객체의 지각도를 보장하고 있다. 즉, 디스플레이 되는 모든 객체의 얼굴을 사용자가 식별할 수 있다.

5. 결론 및 향후 과제

본 논문에서는 효율적인 영상의 UMA 서비스를 위한 PMIT 변환 기법과 인간의 지각적 능력을 영상에 수치적으로 응용한 공간적 해상도 축소 정보를 제안했고, 기존의 방법과 비교해서 그 우수성을 보였다. 그리고 공간적 해상도 정보를 자동적으로 계산하는 기법이나, 색 농도(color depth)에 따른 공간 해상도 값의 차의 특성을 분석해서 비디오나 글자에 자동적으로 적용할 수 있는 알고리즘의 연구가 향후 과제로 남아있다.

6. 참고문헌

- [1] J. R. Smith, R. Mohan, and C. -S. Li, Transcoding Internet Content for Heterogeneous Client Devices, IEEE Intern. Conf. Circuits and Systems, June, 1998.
- [2] J. R. Smith, R. Mohan, and C.-S. Li, Content-based Transcoding of Images in the Internet, IEEE Intern. Conf. Image Processing, Oct., 1998.
- [3] C. -S. Li, R. Mohan, and J. R. Smith, Multimedia content description in the InfoPyramid, IEEE Intern. Conf. Acoustics, Speech and Signal Processing, May, 1998.
- [4] S. Paek and J. R. Smith, Detecting Image Purpose in World-Wide Web Documents, SPIE/IS&T Photonics West, Document Recognition, January, 1998.
- [5] R. Mohan, J. R. Smith and C. -S. Li. Multimedia Content Customization for Universal Access, Multimedia Storage and Archiving Systems, SPIE Vol 3527, Boston, November 1998.
- [6] R. Mohan, J. R. Smith and C. -S. Li. Adapting Multimedia Internet Content for Universal Access, IEEE Transactions on Multimedia, Vol. 1, No. 1, March 1999.
- [7] R. Mohan, J. R. Smith and C.-S. Li. Adapting Content to Client Resources in the Internet, IEEE Intl. Conf on Multimedia Comp. and Systems ICMCS99, Florence, June 1999.
- [8] R. Mohan, J.R. Smith and C-S. Li, Content Adaptation Framework: Bringing the Internet to Information Appliances, Globecom 99, Dec 1999.