

MPEG-4 콘텐츠 저작의 형식 정의와 시스템 개발

김상욱, 김희선, 차경애
경북대학교 컴퓨터학과

e-mail:{swkim, chaka, kimhs}@woorisol.knu.ac.kr

A Formalization and System Development for MPEG-4 Contents Authoring

Sang-Wook Kim, Hee-Sun Kim, Kyung-Ae Cha
Dept of Computer Science, Kyungpook National University

요 약

시각 언어는 인간-컴퓨터 사이의 인터랙션 수단이다. 올바른 인터랙션을 위하여 저작된 콘텐츠와 재생되는 썸의 시맨틱이 인간과 컴퓨터는 물론 인간과 인간 사이에서 일관성과 신뢰성을 가져야 한다. 본 논문에서는 인터랙션의 일관성과 신뢰성을 위한 시각 언어의 시각 문장 집합을 정의하고 형식화한다. 이러한 형식화는 시각 언어의 시각적 저작 내용의 모호성을 제거하여 공동 참여자의 신뢰성을 높인다. 또한 형식 정의에 대하여 저작 콘텐츠와 재생 썸의 동일성을 그래프로 증명한다. 이러한 형식 정의를 MPEG-4 콘텐츠 저작 도구에 적용한다. 즉, 각 객체 사이의 시간관계 정의에 따라 저작하며, 사용자 인터랙션에 따른 썸의 속성의 변화를 효율적으로 지원한다. 저작된 콘텐츠는 썸의 속성 정보에 따라 객체 사이의 동기화를 유지하면서 재생한다.

1. 서론

본 논문은 분산 멀티미디어 환경에서 MPEG-4 콘텐츠를 시각적으로 저작하고 재생하는 시스템을 형식화하고 개발하였다. 이 논문에서는 저작을 위한 시각 문장의 시각원소와 시각 알파벳을 정의하고, 연산 방법은 컴포넌트에 대하여 정의하며, 시맨틱은 속성 심볼 스트링으로 형식화한다.

이러한 형식화는 다양한 저작 시스템에 적용되어 인터랙션의 시맨틱이 일관성과 신뢰성을 가지도록 한다. 그러므로 시각언어가 유한생성집합(Finite Generator Set)에서 어떻게 구성되는가를 보일 뿐만 아니라, 아이콘 객체와 시맨틱 사이의 관계를 형식화하는 프레임틀을 제공한다. 이러한 연구는 멀티미디어 저작도구의 프레임워크를 제공할 뿐만 아니라

VOD 등에 응용할 수 있다.

본 논문의 제 2 절에서는 기본 개념을 기술한다. 제 3 절에서는 콘텐츠 저작과 재생의 형식 모델을 기술하고 객체 사이의 시간관계에 대한 모델을 정의한다.

2. 기본 개념

본 연구에서는 VoD(Video on Demand) 와 같은 비동기형 분산 멀티미디어 모델에 대한 일반적인 시각 인터랙션 모델의 원리와 언어의 구문과 시맨틱을 정의한다. 인간과 컴퓨터의 시각 인터랙션 모델은 그림 1 의 각 HCI 와 같이 인간과 컴퓨터 사이의 아이콘 객체 해석을 위한 번역과 표현 사이의 사이클이다. 그림 1 은 사용자가 시각 인터랙션으로 저

작과 재생을 요청한 멀티미디어 서비스를 서버에서 사용자에게 프리젠테이션 하는 모델이다.

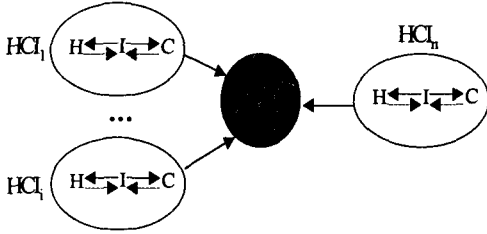


그림 1 멀티미디어 콘텐츠의 저작과 재생 모델

그림 1에서 H 는 인간(사용자) 이며 I 는 해석(번역)하는 에이전트이다. 이 때 t 는 시간으로 사용자가 저작하는 콘텐츠는 각 객체에 대한 저작 시간과 번역에 따른 시간을 나타낸다. C 는 컴퓨터이다.

인터랙션 시스템의 형식화는 아이콘 객체의 시맨틱 분석, 컴퓨터에 의한 아이콘 객체 생성, 인터랙션의 시각 추론을 통합하여 시각언어의 시각원소와 표기 컴포넌트의 형식언어에 대한 특성을 나타낸다. 시각 문장은 다음의 세 컴포넌트로 정의한다.

- (1) 아이콘 객체
- (2) 시맨틱(계산을 수행하는 프로그램)
- (3) 프로그램과 아이콘 객체 컴포넌트 사이의 관계

3. 콘텐츠 저작과 재생의 형식화

아이콘 객체

아이콘 객체 I 는 이차원 스트링 $I : \{1, \dots, r\} \times \{1, \dots, c\} \rightarrow P$ 로서, r 과 c 는 두 정수이며 P 는 알파벳이다. 각 아이콘 객체의 원소는 바로 픽셀인데, 이 픽셀은 (r, c, p)로 구성된다. 여기에서 $r \in \{1, \dots, r_{max}\}$ 는 픽셀의 행이고 $c \in \{1, \dots, c_{max}\}$ 는 픽셀의 열이다. 또한 $p \in P$ 이다.

아이콘 객체 구조

아이콘 객체 구조 s 는 아이콘 객체의 부분집합이다. 아이콘 객체 구조는 특성 함수 $st : \{1, \dots, r_{max}\} \times \{1, \dots, c_{max}\} \rightarrow \{0, 1\}$ 에 의해 규정된다. 또한 $s = \{(r, c, p) \in I \mid st(r, c) = 1\}$ 로 정의된다. 아이

콘 객체 I 내의 모든 아이콘 객체 구조의 집합은 2^I 로 표기된다.

멀티미디어 표현 언어

멀티미디어로 표현되는 언어 ML 은 아이콘 객체의 집합 I 의 부분집합이다. 즉, $ML \subseteq I = \{I \mid \exists r, c \in \mathbb{N}, I : \{1, \dots, r\} \times \{1, \dots, c\} \rightarrow P\}$ 이다. 이차원 스트링으로서의 아이콘 객체의 정의는 여러 방법으로 순서화 된다. 반면에, 일차원 스트링 s 는 $s : \{1, \dots, n\} \rightarrow V$ 로 정의되는데(V 는 알파벳) 선형 순서와 연결된 순서를 가진다.

아이콘 객체의 시맨틱

아이콘 객체 I 의 시맨틱은 함수 $sem : 2^I \rightarrow U$ 이며 U 는 참조 집합이다. 특성 패턴은 $\langle s, u \rangle$ 의 쌍이며, $s \in 2^I, u \in U$, and $u = sem(s)$ 이다. 사용자는 아이콘 객체를 보면서 특성 패턴을 관찰하는데, 각 특성의 구조를 알려 기술한다. 즉, 특성 패턴 각각에 대하여 시맨틱을 제공한다. 또한 아이콘 객체에 대한 해석과 추론을 위해 생성된 쌍을 사용한다. 사용자는 이러한 특성 구조를 응용 영역과 다른 특성(기하학적이거나 위상 등)에서의 시맨틱의 기본으로 삼는다.

아이콘 객체 인터랙션

아이콘 객체 인터랙션에서 U 의 각 원소는 속성 심볼의 스트링이다. 알파벳 V와 값 집합의 집합인 $W = \{P_{a1}, \dots, P_{am}\}$ 가 주어지면, V에서의 속성 심볼은 심볼 $v \in V$ 로 구성되는 (m+1)-튜플이다. 또한 m 은 $p_{ai} \in P_{ai}, 1 \leq m$ 를 특성 짓는다. 각 P_{ai} 가 값의 유한 집합일 때, $W = V \times P_{a1} \times \dots \times P_{am}$ 를 속성 심볼의 알파벳이라 한다.

아이콘 객체의 디스크립션

아이콘 객체의 디스크립션 d 는 알파벳 $W = d : \{1, \dots, g\} \rightarrow W = V \times P_{a1} \times \dots \times P_{am}$ 에 대한 속성 심볼의 스트링이다. 디스크립션 언어 DL 은 가능한 디스크립션의 집합 D의 부분집합으로 $DL \subseteq D = \{d \mid \exists g \in \mathbb{N}, d : \{1, \dots, g\} \rightarrow W\}$ 이다.

인터프리터

아이콘 객체 I 의 해석 int 는 함수 $int : 2^I \rightarrow DL$ 이다. 정의 10은 아이콘 객체를 해석 할 때 사용자의 생각을 구현하는 함수를 형식화하는 것이다. 아

이론 객체는 해석에 의해서 구별된 각 특성 패턴을 나열함으로써 기술된다. 아이콘 객체 디스크립션 d 의 표현 mat 는 함수 $mat:2^d \rightarrow PL$ 인데, 여기에서 2^d 는 d 에서 추출되는 속성 심볼의 모든 집합을 나타낸다. 디스크립션이 주어지면 아이콘 객체를 합성하는 함수이다. 이러한 표현은 원래의 아이콘 객체를 재생할 필요는 없다.

시간 관계 모델

시각 저작을 위한 인터랙션에는 n 개의 미디어 객체에 대한 시간관계가 중요하게 취급된다. 시간관계는 저작/재생에서 동기화의 의미를 가진다.

P 를 $P = \{P_1, P_2, \dots, P_n\}$ 인 n 개의 시간 간격을 가지는 순서 집합이라 하자. 시간관계 α 은 N -시간 관계라 하며 α^i 로 쓴다. 이 때 반드시 다음이 성립하여야 한다.

$$P_i \alpha P, \quad \forall (1 \leq i < n)$$

각 인터벌은 시작 시점(S_i), 기간(D_i), 각 데이터 원소 i 의 종료시점을 가지며, 상대적인 위치와 시간의 종속은 지연(D_{δ}^i)에 의해 얻어진다. 전체적인 기간은 (D_{α}^n)이다. 여러 미디어의 재생에 대한 종료 시간과 시간적인 제약을 가지는 N -시간관계의 성질을 알 수 있다[5]. 멀티미디어 콘텐츠는 이러한 시간관계의 속성을 지니는 멀티미디어 객체의 복합체이다. 그러므로 콘텐츠의 씬은 시간관계 모델에 따른 그래프 모델이다. N -시간관계의 시맨틱과 객체의 동기화 기술을 위해 멀티미디어 객체를 그룹화하고 시간에 관한 인자를 가진다.

터미널 노드 타입은 미디어 타입(텍스트, 이미지, 비디오 등)을 가지고, 동기화 레벨에서는 가장 작은 원소이다. 너터미널 노드는 노드 타입, 자식 노드에의 포인터, 시간 인자(앞으로 지연, D_{δ}^{-i} , 역으로 지연, D_{δ}^{-i} , 전체 시간, D_{α}^n)이 있다. 이러한 인자는 시간에 종속되는 데이터의 검색을 위해 유지된다.

시각 문장

시각 문장 vs 는 $\langle I, d, \langle int, mat \rangle \rangle$ 의 트리플인데, I 는 아이콘 객체, d 는 디스크립션, int 는 해석 함수, mat 는 표현 함수이다. 세 구성 요소를 구별하고 분석하여 번역하는 과정은 다음과 같다.

- (1) 주어진 아이콘 객체를 해석한다.
- (2) 주어진 디스크립션에서 아이콘 객체를 합성한 다.
- (3) 주어진 아이콘 객체와 디스크립션 사이의 관계를 설정한다.
- (4) 시각 언어를 사용한다.

이러한 기본 요소에서 시각 문장을 구성하고, 시각 문장은 시각 언어를 생성한다. 즉, 시각 언어는 시각 문장의 집합이다. 그러므로 시각 언어 VL 은 모든 시각 문장이 완전할 때 완전하다고 한다.

저작과 재생 콘텐츠의 동일성

저작에 의하여 생성되는 콘텐츠에 대한 트리를 $Tr(A)$ 라하고, 재생을 위해 재구성되는 트리를 $Tr(P)$ 라 하자. 이 때 $Tr(A)$ 와 $Tr(P)$ 가 동일하다는 것은 $Tr(A)$ 와 $Tr(P)$ 원소의 수(N)와 대응되는 에지(E), 저작된 정보와 재생되는 정보(Inf)가 일치하면 $Tr(A) = Tr(P)$ 이다.

4. MPEG-4 콘텐츠 저작과 재생에의 적용

위에서 언급된 콘텐츠 저작에 대한 형식 정의는 다양한 멀티미디어 저작 시스템에 적용할 수 있다. 이 절에서는 MPEG-4의 콘텐츠 저작을 위한 BIFS를 생성하는 시스템에 적용한다.

그림 2는 인터페이스에서 시각적으로 저작된 씬이 재생기를 통하여 재생되는 화면이다.

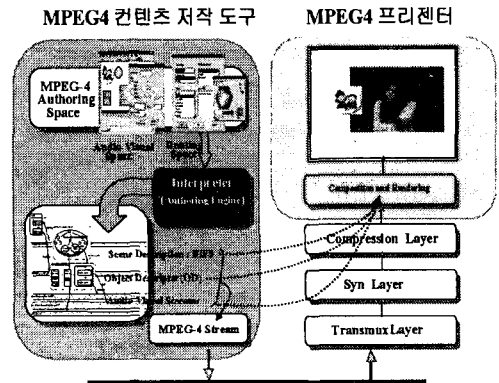


그림 2. MPEG-4 저작/재생 시스템 구조

그림 2에서 인터페이스는 시청각 객체를 이용하여 씬을 저작하도록 아이콘 기반 저작 환경을 제공한다. 사용자는 각 객체에 대한 속성과 자신의 재생 시간을 결정하는 타임 슬라이드를 가진다. 사용자로 하여금 특정 객체의 반응을 대화형으로 입력하도록 하여 라우팅 정보를 쉽게 저작한다. MPEG-4 씬 생성 모듈은 사용자 인터페이스로부터 저작된 씬에 대한 시청각 정보에 의해 MPEG-4 씬을 씬 디스크립션으로 변환한다. 씬 구조는 저작된 시간관계 모델에 의한 객체사이의 동기화 정보를 그룹화한다. OD 생성기는 씬 그래프 내에 객체의 부피가 큰 객체의 객체 디스크립터를 생성한다. 라우트 정보 생성기는 시공간적 관계를 표현한 생성된 씬 그래프를 변환하여 저장한다.

MPEG-4 기반의 씬 구성과 재생은 DMIF와 연동하여 채널을 요청 및 생성함으로써 이루어진다. 생성된 채널로부터 스트림을 전송받아 파싱한 후, 컴포지터가 씬을 구성하면 프리젠테이션이 랜더링한다.

그림 2의 MPEG-4 씬 저작 도구는 윈도우 NT/98 환경에서 Visual C++ 로 구현하였다.

5. 결론

이 논문에서는 멀티미디어 콘텐츠를 시각적으로 저작하고 재생하는 과정을 각 미디어에 대한 시각적인 아이콘 객체와 시맨틱 사이의 관계를 형식화하였으며 MPEG-4 콘텐츠 저작 시스템에 활용하였다[6,7]. 이러한 형식화는 사용자와 컴퓨터 사이의 인터랙션에 대하여 시맨틱이 일관성과 신뢰성을 가지므로 시각언어가 유한생성집합(Finite Generator Set)에서 어떻게 구성되는가를 보일 뿐만 아니라 아이콘 객체와 시맨틱 사이의 관계를 형식화하는 프레임워크를 제공한다.

이 논문에서의 형식화 모델을 MPEG-4 콘텐츠의 시각적인 저작 도구와 재생기에 적용하여 MPEG-4 저작/재생기를 개발하였다.

본 연구는 VOD 등에 효율적으로 활용할 수 있다. 또한 다양한 대화형 멀티미디어 응용 프로그램을 지원하는 프레임워크가 될 수 있다. 앞으로의 연구는 시각 속성과 재생에 대한 좀 더 구체적인 형식화와 3차원 저작과 랜더링을 추가한다.

참고문헌

- [1] 김상욱, 김희선, 차경애, 권순동, 배수영, 최상길, 정영우, "MPEG-4 기반 멀티미디어 재생 엔진", '98 가을 학술 논문 발표집, 한국정보과학회, 제 25 권, 제 2 호, 1998.
- [2] Part 6:Delivery Multimedia Integration Framework, ISO/IEC CD 14496-6, ISO/IEC JTC 1/SC 29 N1906, 31. 10. 1997.
- [3] Atul Puri and Alexandros Eleftheriadis, "MPEG-4: An object-based multimedia coding standard supporting mobile applications," Mobile Networks and Applications, 1998.
- [4] ISO/IEC FCD 14496-1, Information technology Generic Coding of A/V Objects Part 1 :Systems, ISO/IEC/JTC1/SC29/WG11, 1998, 5.
- [5] Thomas D.C. Little and Arif Ghafoor, "Interval-Based Conceptual Models for Time-Dependent Multimedia Data," IEEE Tr. on Knowledge and Data Engineering, Vol. 5, No. 4. Aug. 1993.
- [6] K.Marriot, Visual Language Theory, A-W, 1999.
- [7] K. Cha, H. Kim and S. Kim, MPEG-4 Contents Authoring Tool and Player, HCI 2000, 학술대회발표논문집, pp. 27-30, 2000.