Multimedia Content Description Interface(MPEG-7)에서의 Fuzzy화 가능성

○

이상영, 김영택

경성대학교 전자계산학과

Feasibility on Fuzzy Inferenced Multimedia Content Description Interface(MPEG-7)

○

Sang-young Lee, young-taek Kim

Dept. of Computer Science, Kyungsung University

요약

MPEG-7의 Standard Description을 위한 여러 가지 요소에 대한 정의가 여러 연구에서 약간의 차이를 주고 있지만 전체적인 방향은 형성되어 지고 있는 실정이다. 본 연구에서는 그 요소들에서 지능형 멀티미디어 탐색엔진에서 high level linguistic query를 처리할 수 있는 fuzzy inferenced database 기능을 MPEG-7과 연계할 수 있는가를 연구하였다.

Ⅰ. 서론

현재 전세계적으로 멀티미디어의 컨텐트(content)화에 대한 관심으로 고조되고 있다. 전세계의 멀티미디어 산업의 척도 역시 얼마나 많은 멀티미디어 컨텐트가 제작되는 가로 판단되어 질 수 있다. 그러나 무엇보다도 MPEG(Moving Picture Experts Group)은 멀티미디어 컨텐트의 저장에 있어서 가장 중요한 역할을 하고 있다. 이는 동영상이 가장 원초적인 멀티미디어의 표본이라고 볼 수 있기 때문이다.

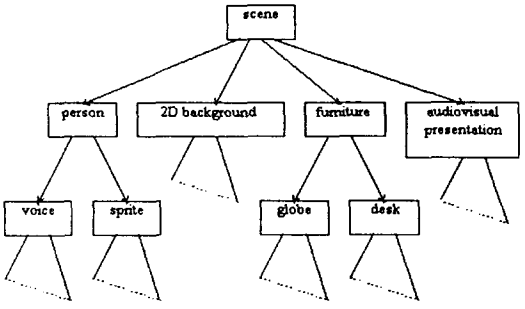
MPEG-1, 2에서 보여졌던 압축에 관한 표준은 이제 MPEG-4에서 AV 자료(Audio-Visual Information)를 디지틀화(digitalization)시키는데 그치지 않고 이를 객체 단위로 부호화 하는데 이르렀다. MPEG-4의 표준화가 완성된 상태에서 'Multimedia Content Description Interface'라 불려지는 MPEG-7에서는 컨텐트의 객체들을 description scheme에 의해 계층적(hierarchical)으로 구조화 시켜 표준화 하려한다.[1] 이렇게 표준화한 description을 이용해 정보검색에 적용될 수 있도록 한 것이다. 현재 상용으로 널리 이용되고 있는 멀티미디어 데이터베이스에서 주어진 정지영상(still picture), 동영상(moving picture), 그리고 음성(voice) 등을 데이터베이스화한 후 이를 query할 수 있다.

2000년에 정식으로 표준화 작업이 완료될 MPEG-7에서 멀티미디어 객체의 description에 fuzzy의 소속함수를 추가하고 fuzzy에 의한 처리를 제안하고자 한다.

Ⅱ. 본론

1. Multimedia Content Description Interface - MPEG-7

이미 표준화되어 Digital TV, HDTV, DVD, Video CD등에서 실제로 사용되고 있는 MPEG-1, 2는 영상을 컨텐트화 시키는데 있어서 얼마나 적은 공간에 저장하는가가 주 관심의 대상이었다. 또한 지금도 동영상(moving picture)뿐만 아니라 음성이나 일반 이진 파일(binary file)에 대한 압축(compression)의 연구는 계속 진행되고 있다. 그러나 비디오에 대한 압축에 관심을 가지는 것이 아니라 멀티미디어 컨텐트를 객체화시키는 시도가 MPEG-4에서 이루어 졌다. MPEG-4의 AVO(Audiovisual Object)가 그것인데, [그림 1]은 MPEG-4의 scene을 계층적인 구조로 표현한 예이다.[2]



[그림 1] Logical structure of a scene

현재의 많은 정보는 audio-visual information에 의해 표현되어지며 또한 사용되고 있다. 이러한 정보는 일반 텍스트 정보(textual information)와 같이 처리하는 것이 단순하지 않다. 우리는 일반적으로 인터넷에서 문서를 검색하거나 데이터 베이스에서 문자열을 검색하기 위해 단순히 같은 문자열을 매칭하면 된다. 그러나 정지영상(still picture), 3D 모델링(3D Models), 그래픽스(graphics), 소리(audio), 음성(speech), 동영상(video) 자료(보통, scenario, 또는 composition information으로 통칭됨)는 이와 같은 textual information과 달리 정보를 판별(identifying)하거나 검색(searching)은 용이하지 않을뿐더러 상당히 복잡하다. 간단히 예를 들어 비디오에서 자신이 원하는 영상을 찾거나 좋아하는 장면으로 이동하는 것은 불가능하다. 앞으로 감기(forward)또는 되감기(rewind)는 상당히 비효율적(inefficiency)이라는 사실에 주목해야만 한다. 현재 상업적으로 흔히 사용되고 있는 데이터베이스에서는 multimedia 데이터를 판별(identify)하거나 검색(query)하기 위한 기능을 지원하고 있다. MPEG-7에서는 이러한 멀티미디어 데이터를 처리하기 위해 'Multimedia Content Description'을 표준화하고자 하는 것이다. 이전의 표준인 analogue, PCM, MPEG-1, 2, 그리고 MPEG-4와 전혀 무관한 것이 아니며, 특히 MPEG-4의 AVO(Audio-Visual Object)과 아주 많은 관련성을 지니고 있다.[3]

MPEG-7에서 정의되어 사용되는 terminology는 다음과 같다.[5]

◆Data MPEG-7에서 사용되는 AV information, 예를 들어 MPEG-4 stream, video tape, 또는 book이 될 수 있다.

◆Feature data에서 객체의 특징을 구분할 수 있는 것이다. 예를 들어 객체의 색, 크기, 이동 경로 등이 될 수 있다.

◆Descriptor 객체의 feature을 묘사하기 위한 표현 값이다

◆Description Scheme(DS) descriptor의 구조(structure)나 구문(semantics)그리고 content의 관계를 기술한다

◆Description 데이터와 DS로 구성되는 개체, descriptor를 설명하는 개체이다.

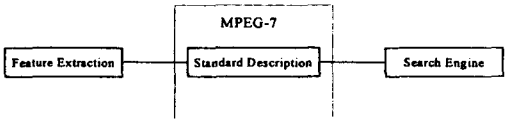
◆Coded Description 인덱싱과 효과적인 저장(storage)과 전송(transmission)을 위해 description을 압축한다.

◆Description Definition Language(DDL) 새로운 DS를 생성하고 이미 존재하는 DS를 확장하기 위해 필요하다.

MPEG-7의 content description을 다루기 위해서는

◆Feature Extraction
◆Content Description
◆Retrieval Engine

이 필요한데, 실제로 MPEG-7에서 표준화될 대상은 content의 description이다.[3]



[그림 2] Relationship of three aspects

일반 상용 멀티미디어 데이터베이스는 feature extraction과 description, 그리고 search engine(query processing)을 모두 가지고 있다. 그러나 MPEG-7에서 표준화의 대상은 멀티미디어 content의 description에만 한정된다.

MPEG-7에서는 AVO를 describe하기 위해 두 개의 level로 나누고 있다. low abstraction level과 high abstraction level(semantic level)로 나눌 수 있는데, low level은 scene내에 존재하는 물체들의 모양(shape), 크기(size), 색감(texture), 색깔(color), 이동(movement, trajectory), 그리고 위치(position)등 객체의 물리적인(physical) 내용을 담고 있다. 그리고 semantic level에 의한 객체의 표현은 언어적인 정보에 의한 것이다. 예를 들어 '두 남녀가 손을 잡고 걸어가는데 축구공이 두 사람의 머리를 향해 날아오고 있다.'라는 식의 정보로 한 장면을 표현할 수 있다. 이와 같은 표현은 실제로 자동화(automatically)에 의해 효과적으로 생성되기는 어렵다. 이러한 정보의 추출은 대화식 추출 방법(interactive extraction method)에 의해 가능하다. 그러나 low level에서의 표현은 충분히 자동화될 여지가 충분하다.

2. Description Scheme

Description Scheme(이하, DS)의 정의나 그의 요소들은 적용 대상이나 저자에 따라서 각각 달리 정의된다. 'MPEG-7 Requirement'[1]에서는 이것에 필요한 요소들을 정의해 두었다. visual data의 경우 필요사항을 아래에 나열하였다.

- Type of feature color, visual object, texture, sketch, shape, still and moving images(예, Thumbnails), volume, spatial relations, motion, deformation(예, 객체의 겹침), models(예, MPEG-4 SNHC)
- Data visualization using the description
- Visual data formats MPEG-1, MPEG-2, 또는 MPEG-4, 기타 analogue video, 그리고 정지영상인 JPEG등이 될 수 있으며 graphics(CAD), 3D modeling(VRML), 기타 시각 자료이다.
- Visual data class natural video, still picture, graphics, animation(2-D), 3-D models, 기타 composition information을 정의해 두었다. MPEG-4에서는 natural video, still picture, graphics, composition information가 이미 적용될 수 있다.

위에서 본 바와 같은 조건에 따라 MPEG-97에서 'Pascal Faudemay', 와 'Yong Rui'는 DS을 다음과 같은 요소들로 정의될 수 있음을 제안하였다.[8]

$O = O(D, S, F, R, A, W)$

- D는 데이터 소스로서 MPEG movie나 analog movie등을 의미한다.
- S = $\{s_i\}$ 는 "subobject"라는 용어로 객체 구조이다. 이것은 time interval, shot, 프레임의 영역(region of frame)등이 될 수 있다.
- F = $\{f_i\}$는 객체와 관련된 feature이다. 색깔(color), 색감(texture), 모양(shape)등의 비디오의 움직임에 관한 요소들이 이에 속한다.
- R = $\{r_i\}$은 주어진 feature f_i에 대한 표현이다. 예를 들어 색깔의 특징은 색깔의 히스토그램(color histogram)과 색의 이동(color movement)는 색깔 특징에 대한 표현이다.
- A = $\{a_k\}$는 데이터와 feature 또는 feature type의 관계, feature과 feature type의 관계 몇 개의 데이터와의 연관성을 나타낸다. 예를 들어, "production date"는 어떤 feature type과 1:1로 대응될 수 있다.
- W = $\{w_1\}$은 데이터와 컨텐트에 존재하는 부가적인 정보이다. 이 정보는 description author또는 description supplier, 또는 description author의 어떤 부가적인 정보이다.

또한 Illinois 대학의 Multimedia Analysis and Retrieval System(MARS)에서는 멀티미디어 객체를

$O_M = O_M(D, F, R, M, V)$

로 정의하였다.[7]

여기서

◆D는 객체의 raw data이다. 예를 들어, JPEG 이미지 또는 MPEG 비디오가 될 수 있다.

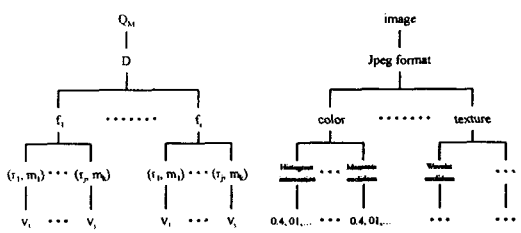
◆F = $\{f_i\}$는 객체와 관련된 특징, 즉, 비디오의 움직임에 대한 요소는 색깔이라든지, 질감, 그리고 이미지의 모양을 들 수 있다.

◆R = $\{r_j\}$은 주어진 Feature f_i 에 대한 표현이다. 색깔의 히스토그램(color histogram)과 색깔의 모멘트(color moment)는 색깔에 대한 특징(color feature)이다.

◆M = $\{m_k\}$은 feature f_i를 파악하기 위해 m_k는 r_j와 결합된다. 예를 들어, color feature(f_i)을 파악하기 위해 히스토그램 교차(histogram intersection)(m_k)는 색깔 히스토그램(color histogram)(r_j)와 함께 사용된다.

◆V = $\{v_j\}$는 R의 실제 값이다. 각각의 r_j에 대해 그 표현에 대한 실제 값이 저장되는 v_j이 있다.

이러한 객체 모델은 계층적으로 가시화할 수 있는데, [그림 3]는 계층적 모델을 가시화한 것과 그에 대한 예를 보인 것이다.



[그림 3]Multimedia object model

또한 'Yong Rui', 'Thomas S. Huang', 그리고 'Shih-Fu Chang'은 멀티미디어 객체를 $O=O(D,F,R)$ 에 의해서 정의하였다.[4]

3. Rational consensus model - Group decision making problem

제안되고 있는 MPEG-7의 description 요소에서 $R=\{r_j\}$ (set of representation for a given f_i)과 $M=\{m_k\}$ (set of similarity measures) 두 요소가 feature f_i에 대한 표현과 그 표현의 인지 정도를 결정하는 중요한 사항이 될 수 있다. 여기서 multiple-element F set이 multimedia object에서 color, texture, shape, layout, motion parameters등 각각 여러 가지 (r_j,m_k) combination을 이루면서 어떤 인지 문제에서 상황에 따라 다른 종류의 결론을 얻을 수 있는 perspection criteria를 구성하고 있다. 이런 방법의 modeling에서 (F, R, M) set가 database에서 multimedia object로 저장되고 또한 query Q_M이 해당되는 O_M을 similarity sim(Q_M, O_M)의 criteria를 지원할 수 있다면 적당한 fuzzy inferenced query engine을 형성할 수 있는 model로 확장할 수가 있을 것이다.

그 이유는 (r_j, m_k)combination이 각 다른 사람(expert)들의, 혹은 같은 사람일지라도 각기 다른 상황(circumstance)에 따라 다른 perception criteria를 어느 한 feature에 적용할 수 있을 것이고 이런 관점에서의 의사결정(decision making) 문제가 유연성(flexibility)있는 결과를 가져오기 위해서는 여러 명의 expert들의 공통된 의견(consensus)으로서 결론을 도출할 수 있는 heterogeneous group decision making problem으로서 해결책을 찾아야하고 인간의 사고 방식과 일치할 수 있는 fuzzy multi-person decision making model을 정의하여 수행하는 것이 바람직하게 보여진다.

여기서 이런 방법의 표현을 rational consensus model이라고 부를 것이다.[9]

이런 표현을 수행하기 위해서 higher level representation 요소이면서 fuzzy logical inference에 필수적인 linguistic preference relation을 database에 사용할 수 있어야 하고 expert의 opinion으로서 consensus solution을 얻는데 필요한 단서로 작용해야만 한다.

4. Measuring Process

위 장에서의 feature들 중 linguistic fuzzy logical expression을 가지고 heterogeneous set of experting $E=(e_1,e_2,...,e_m)$ 을 구성하여 요소로서 사용할 수 있다면 각 $e_k \in E$들에는 각각 정의된 중요도(importance degree) 소속값 $\mu_{e_k}(k)$값이 설정되어야 한다.

여기서 $\mu_{e_k}(k)$들이 group expert들로부터 multiple

decision making model을 구성할 수 있고 그 소속 값들이 fuzzy inference를 가능하게 하면서 MPEG-7과 같이 higher level description을 이루어 줄 수 있다.

rational consensus model을 구성하기 위해 필요한 처리과정으로서 두 가지 사항을 든다면 F. Herrera에 의해 제안된 consensus measuring process와 consistency measuring process가 적당한 수단이 될 수 있다. 여기서 consensus measuring은

- counting process
- coincidence process
- computing process

로 처리될 수 있고 그 각각은 expert의 개인 의견을 count하고 또 coincidence degree를 계산하여 대다수의견(majority opinion)의 preference value를 계산함으로서 진행할 수 있다. 마지막 step인 computing process에서 linguistic consensus measure가 계산되어져서 linguistic consensus degree를 얻을 수 있다. 제안된 consensus computing 과정에는 크게 3가지 레벨로 구성할 수 있고 그 각각은 preference linguistic, alternative linguistic, relation linguistic consensus로 이루어 질 수 있다.

여기서 만들어진 consensus measure가 consistency measure process 과정의 결과와 함께 moderator가 recommendation을 만들어 낼 수 있게 한다.

물론 linguistic consensus degree 와 linguistic distance 두 가지 type의 measure가 사용 될 수 있다.

Ⅲ. 결론

multimedia content description interface에 필요한 각종 요소들이 현재까지 확실한 standardization이 결정되지 않은 상태에서 지능형 search engine을 위해 fuzzy relational database의 필요성이 있게 되고 experting에서의 각종 opinion들의 consensus과정이 연구되어져야 한다.

그렇다면 consensus 도출에 따른 fuzzy inference 기능과 content description이 decision making을 가능케 하고 또한 flexible한 multimedia database가 구성될 수 있기 때문에 본 연구에서는 rational consensus measure에 대한 사용가능성을 제안하는 것이고 차후 실제적인 실험 연구에서 MPEG-7의 operation을 현실화시키고자 한다.

참고문헌

[1]MPEG-7 Requirements, ISO/IEC JTC1/SC29/WG11/N2208, MPEG98, March 1998

[2]Overview of the MPEG-4 Version 1 Standard, ISO/IEC JTC1/SC29/WG11/N1909, MPEG97, Oct 1997

[3]MPEG-7 Context and Objective, ISO/IEC JTC1/SC29/WG11/N2207, MPEG98, March 1998

[4]Digital Image/Video Library and MPEG-7 : Standardization and Research Issues, Yong Rui, Thomas S. Huang, Shih-Fu Chang

[5]Suggestions to the Draft of MPEG-7 Requirements, ISO/IEC JTC1/SC29/WG11/M3107, MPEG97

[6]MARS and Its Applications to MPEG-7, ISO/IEC JTC1/SC29/WG11, MPEG97, July 1997

[7]A Relevance Feedback Architecture for Content-based Multimedia Information Retrieval Systems, Yong Rui, Thomas S. Huang, Sharad Mehrotra, and Michael Ortega

[8]contribution to MPEG-7 Proposal Package Description(PPD), ISO/IEC JTC1/SC29/WG11, MPEG97, January 1998

[9]A rational consensus model in group decision making using linguistic assessments, F. Herrera/E.Herrera-Viedma/J.L. Verdegay, Fuzzy Sets and Systems 88, 1997