

시공간 관계를 지원하는 비디오 모델링

복경수^o 유재수

충북대학교 정보통신공학과 및 컴퓨터·정보통신연구소

ksbok@netdb.chungbuk.ac.kr

yjs@cbucc.chungbuk.ac.kr

Video Modeling Supporting Spatio-Temporal Relationship

Kyoung Soo Bok^o Jae Soo Yoo

Dept. of Computer & Communication, Chungbuk National University

요 약

최근 컴퓨터 응용 기술의 발달로 비디오 데이터에 대한 처리의 필요성이 증가하고 있다. 비정형화된 비디오에 대한 검색을 효율적으로 처리하기 위해서는 비디오의 논리적 구조와 의미적 내용을 표현할 수 있는 비디오 모델링 기법이 필요하다. 본 논문에서는 비디오의 논리적 구조는 물론 비디오 내에 포함된 의미적 내용을 표현하기 위한 비디오 모델링 방법을 제안한다. 제안하는 모델링 기법은 의미적 내용을 효과적으로 표현하지 못하는 기존의 구조 모델링의 문제점을 해결하고 의미적인 내용들간의 시공간적 관계를 정의한다. 또한 시공간적 관계를 통한 의미적 내용에 대한 검색을 효과적으로 수행할 수 있도록 한다.

1. 서론

최근 컴퓨터 응용 기술의 발달로 비디오 데이터에 대한 사용이 증가하고 있다. 이와 함께 비디오 검색을 위한 많은 시스템들이 개발되었다. 일반적으로 비디오 데이터는 대용량이며 비정형적인 데이터 특성을 포함하고 있다. 또한 비디오는 시청각(audio-visual) 특성은 물론 객체와 사건과 같은 비디오의 의미적 내용들 사이에 시공간(spatio-temporal) 특성을 포함하고 있다.

이러한 비디오에 대한 효과적인 접근을 수행하기 위해서는 비디오를 논리적인 구조로 표현하거나 비디오 내에 포함된 의미적 내용과 시각적인 내용을 표현하기 위한 비디오 모델링 기법이 필요하다. 또한 비디오의 논리적 계층 구조와 의미적 내용들 사이의 시공간적인 관계를 정의할 수 있어야 한다.

기존의 모델링 기법들은 구조적 계층 내에 시각적 특징 또는 의미적 내용을 표현하기 때문에 내용 정보들이 특정 구조 계층 내에서만 의미를 갖는다. 또한 의미적 내용에 대한 시간적인 존속 기간을 효과적으로 표현하지 못한다는 문제점이 있었다.

이러한 문제점을 해결하기 위해 본 논문에서는 비디오의 구조 모델과 내용 모델로 계층화된 모델링 기법을 제안한다. 또한 논리적 구조, 시각적 내용 그리고 의미적 내용에 대한 시공간 관계를 정의한다. 제안하는 비디오 모델링 기법은 원시데이터 질의, 구조 질의, 유사도 질의 및 시공간 관계에 기반한 내용 질의 등을 지원한다.

이 논문의 구성은 다음과 같다. 제2장에서는 기존에 제시된 비디오 모델링 방법에 대해서 기술하고 제3장에서는 제안하는 비디오 모델링 기법과 시공간 관계에 대해 기술한다. 제4장에서는 제공되는 질의 타입에 대해 기술하고 5장에서는 결론과 향후 연구 방향을 제시한다.

* 본 연구는 과학재단 특정기초과제(과제번호 : 1999

-1-303-007-3) 연구비 지원에 의하여 수행되었음

2. 관련 연구

비디오 모델링 기법은 크게 구조 모델링과 내용 모델링으로 구분된다. 구조 모델링은 논리적인 구조 계층에 의해 비디오를 구조화하는 모델링 방법이다. 이에 반해 내용 모델링은 비디오 내에 포함된 의미적 내용과 시각적 내용을 이용하여 비디오의 내용을 모델링한다.

[1]에서는 비디오의 특성을 7개의 관점에서 수정이 없이 사용될 수 있는 일반화된 메타데이터 모델 GMM을 제안하였다. GMM은 5 계층의 논리적인 계층 구조에 의해 비디오 데이터베이스에 존재하는 비디오 데이터와 메타데이터를 공유하여 효과적인 검색을 수행할 수 있도록 하였다. [2]에서는 비디오의 논리적 계층 내에 사건과 객체를 표현하는 모델링 기법을 제안하였다. 또한 각 계층 구조에 대한 시간 간격을 표현하여 비디오 데이터의 구조에 따라 3가지 시간 관계성을 정의하였다. [3]에서는 비디오에 대한 주석기반 검색과 유사성 검색을 두 개의 계층으로 이루어진 통합객체지향 모델 THOMM을 제안하였다. THOMM은 구조적 계층 간의 시간적 관계는 물론 씬 내에 존재하는 객체에 대한 정보를 이용한 유사도 검색을 지원한다.

그러나 [1, 2, 3]에서 제안된 모델링은 비디오의 의미적 내용들이 구조적 계층 내에 포함되어 있기 때문에 시간적인 존속기간을 정확히 표현하지 못한다. 또한 [1, 3]에서는 비디오에 나타난 사건에 시간 관계와 시간의 변화에 따른 객체의 동작 변화를 표현하지 못한다. [2]에서는 객체에 대한 시간 관계는 효과적으로 표현하지만 의미적 내용들에 대한 공간 관계를 표현하지 못하는 문제점이 있다.

3. 제안하는 비디오 모델링 기법

3.1 비디오 모델링

제안하는 비디오 모델링은 [그림 1]과 같이 비디오의 의미적 단위에 따라 계층화되어 있다. 비디오 모델링의 최하위에는 비디오의 상영 시간, 저장 방식, 제목, 제작

자, 출연진과 같이 비디오 자체가 갖고 있는 가공되지 않은 정보를 표현하는 원시 데이터 계층이 존재한다. 원시 데이터 계층의 상위에는 비디오의 논리적인 구조 정보를 표현하는 구조 모델 계층과 비디오 내에 포함된 내용을 표현하는 내용 모델 계층이 존재한다.

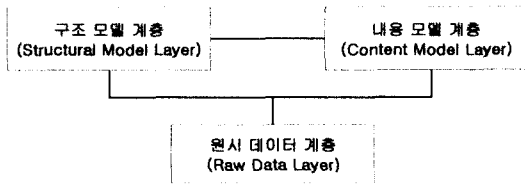


그림 2 비디오 모델링의 구조

비디오의 구조 모델 계층은 비디오에 대한 논리적 구조를 표현하는 모델 계층으로 비디오(video), 시퀀스(sequence), 씬(scene), 샷(shot)으로 이루어져 있다. 각 논리적 구조 계층은 (*sid*, *T*, *E*, *N*)으로 이루어져 있다. 이때, *sid*는 구조 계층의 식별자, *T*는 각 논리적 구조 단위의 시작 프레임 *frame_S*과 끝 프레임을 나타내는 *frame_E*에 의해 논리적 구조 단위의 존속 기간을 나타내는 [*frame_S*, *frame_E*], *E*는 구조 계층 내에서 발생한 카메라 효과 또는 특수 효과, *N*은 구조 계층에 부여된 주석 정보이다.

이에 반해 비디오 내용 모델 계층은 구조적 계층 정보와는 독립적으로 [그림 2]와 같이 시각적 특징과 의미적 내용을 표현한다.

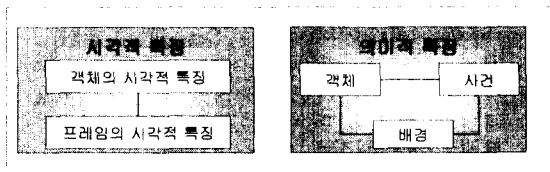


그림 3 내용 모델 계층

비디오에 포함된 시각적 특징을 표현하는 시각적 특징 계층은 하나의 프레임에 대한 시각적 특징과 프레임 내에 포함된 객체의 시각적 특징으로 이루어져 있다. 프레임의 시각적 특징은 샷에서 추출된 키프레임에 대한 시각적 특징으로 표현한다. 또한 프레임 내에는 하나 이상의 객체들이 존재한다. 객체의 시각적 특징은 특정 시간에 나타난 객체들 사이의 공간적 관계를 표현하기 위해 객체들에 대한 공간적인 위치를 MBR(Minimum Boundary Rectangular)에 의해 표현하고 해당 객체의 시각적 특징 정보를 표현한다. 따라서 시각적 특징은 (*frameid*, *C*, *F*, *OF*)과 같이 나타낸다. 이때, *frameid*는 키프레임 식별자, *C*는 시각적 특징의 개수, *F*는 프레임에 대한 특징의 정보 리스트, *OF*는 프레임에 나타난 객체의 MBR 정보와 시각적 특징을 나타낸다. 프레임에 나타난 객체의 시각적 또는 공간적 특징을 나타내는 *OF*는 $\langle\langle oid_1, MBR_1, F_1 \rangle, \dots, \langle oid_n, MBR_n, F_n \rangle\rangle$ 와 같이 객체의 식별자 *oid_i*, 객체의 MBR을 나타내는 *MBR_i*, 그리고 객체의 시각적 특징을 나타내는 *F_i*에 의해 나타낸다.

내용 모델의 의미적 특징은 비디오에 나타난 객체와 사건 또는 배경과 같이 비디오의 의미적 내용을 표현한다. 의미적 내용은 다른 의미적 내용과 서로 연관되어 비디오에 나타난다. 따라서 의미적 특징 내에서는 이러한 의미적 내용들 사이의 시공간적 관계를 표현해야 한다.

객체(object)는 시간의 변화에 따라 특정한 움직임에 포함하고 있으며 특정 시간 간격을 갖고 연속된 동작(action)을 수행한다. 따라서 객체는 (*oid*, *Name*, *T*, *H*, *A*)과 같이 표현한다. 이때 *oid*는 객체의 식별자, *N*은 객체의 이름, *T*는 객체가 나타나는 시간 간격, *H*는 시간의 변화에 따른 객체의 위치변화, *A*는 객체가 수행하는 동작의 변화를 나타낸다. 시간에 따른 객체의 위치변화 *H*는 $\langle(D_1, S_1, T_1), (D_2, S_2, T_2), \dots, (D_n, S_n, T_n)\rangle$ 로 표현된다. *D_i*는 객체의 움직임 방향을 나타내며, *S_i*는 *D_i* 방향으로 움직인 거리, *T_i*는 *D_i* 방향으로 움직인 시간 간격을 나타낸다. 객체가 시간의 변화에 따라 수행하는 동작 *A*는 $\langle(A_1, T_1), (A_2, T_2), \dots, (A_m, T_m)\rangle$ 로 표현된다. *A_i*는 객체가 수행하는 동작의 이름, *T_i*는 객체가 *A_i*를 수행하는 시간 간격을 나타낸다.

사건(event)은 특정 객체와 연관되어 수행된다. 따라서 사건은 (*eid*, *name*, *O*, *T*)와 같이 표현된다. 이때, *eid*는 사건에 부여된 식별자, *name*은 시간의 이름, *O*는 특정 사건과 관련된 객체의 리스트, *T*는 사건이 나타나는 시간 간격을 나타낸다.

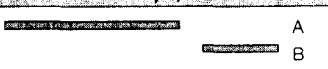






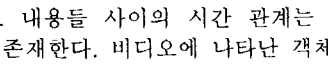
배경(background)은 객체 또는 사건이 발생하는 지역 또는 상황에 관한 정보를 표현한다. 배경은 (*bid*, *name*, *T*)에 의해 나타낸다. 이때, *bid*는 배경에 부여된 식별자, *name*은 배경의 이름, *T*는 배경이 나타나는 시간 간격을 나타낸다.

3.2 시공간 관계

비디오에 포함된 객체와 사건 또는 구조적 계층 간에는 시간적인 이전(before), 이후(after), 겹침(overlap)과 같은 시간 관계의 표현이 필요하다. 뿐만 아니라, 비디오에 포함된 객체들 사이의 공간 관계 또는 시간의 변화에 따라 움직이는 동적 객체의 궤적(trajjectory)에 대한 공간적 표현이 필요하다.

비디오 데이터에서는 논리적 계층 구조 사이의 시간 관계, 의미적 내용들 사이의 시간 관계 그리고 논리적 계층 구조와 의미적 내용들 사이에 시간 관계가 존재한다. 시간 관계를 통한 질의를 가능하게 하기 위해서는 모델링 내에 이러한 시간 관계를 표현하고 정의할 수 있어야 한다. [4]에서 멀티미디어 데이터에서 사용할 수 있는 13가지 시간 관계를 정의하였다. [4]에서 제공되는 시간 관계는 7가지 기본 관계에 6가지의 역관계로 이루어져 있다. 그러나 6가지 역관계는 before, meets, overlap, finishes, during, starts에 대한 상대적 위치이기 때문에 실제적인 관계는 7가지 기본 관계에 의해 표현된다. 본 논문에서는 [표 1]과 같이 [4]에서 제공되는 7가지 기본 관계에 after 관계를 추가하여 8가지 시간 관계를 사용한다. 이를 통해 보다 다양한 시간 관계를 정의할 수 있다.

표 1 시간 관계

관계	의미
A before B	 A B
A after B	 A B
A meets B	 A B
A overlap B	 A B
A during B	 A B
A starts B	 A B
A finishes B	 A B
A equals B	 A B

비디오에는 비디오 내용들 사이의 시간 관계는 물론 다양한 공간 관계가 존재한다. 비디오에 나타난 객체 O_i 에 대한 MBR은 (x_i, y_i, c_i) 로 표현한다. $x_i = [x_s, x_e]$ 와 $y_i = [y_s, y_e]$ 는 객체에 대한 x축과 y축에 범위를 나타내며 c_i 는 x축과 y축의 중심을 나타낸다. 이러한 MBR을 이용하여 객체들 간의 공간 관계를 계산할 수 있다. 비디오 내용들 사이의 공간 관계는 동일한 시간에 나타난 객체들 사이의 공간 관계, 시간의 흐름에 따른 하나의 움직임 객체의 공간 관계들이 존재한다. [5]에서는 동적 객체에 대한 공간 관계를 표현하기 위해 방향 관계와 위상 관계를 정의하였다. 본 논문에서는 공간 관계를 [표 2]와 같이 객체의 방향 관계, 위상 관계 그리고 거리 관계를 정의한다. 방향 관계와 위상 관계는 [5]에서 유사하며 거리 관계는 객체들 사이의 거리가 임계치보다 적을 때에는 near, 그렇지 않으면 far로 정의한다.

표 2 공간 관계

관계	의미
방향 관계	South, North, West, East, Northwest, Northeast, Southwest, Southeast, Equal, Left, Right, Below, Above
위상 관계	Equal, Inside, Contain, Cover, Covered By, Overlap, Touch, Disjoint
거리 관계	far, near

4. 제공되는 질의 타입

본 논문에서 제안하는 비디오 모델링 기법은 원시 데이터와 논리적인 구조에 대한 텍스트 검색, 시각적 특징에 대한 유사도 검색 그리고 시공간 관계에 기반한 내용 검색이 가능하다. 제안하는 비디오 모델링에서 제공되는

질의 타입은 다음과 같다.

(1) 원시데이터 질의

제안하는 모델링의 원시데이터 계층에 존재하는 비디오 자체가 갖고 있는 제목, 저장방식, 출연진, 제작자 등의 정보를 이용하여 원시데이터에 대한 질의가 가능하다.

(2) 구조 질의

비디오의 논리적 구조에 대한 질의는 샷, 씬, 시퀀스에 부여된 주석 정보와 특수 효과 또는 카메라 효과를 이용하여 논리적 구조 검색은 물론 논리적 구조 단위에 표현된 프레임 간격을 통해 시간 관계를 이용한 질의가 가능하다.

(3) 유사도 질의

제안하는 모델링의 시각적 특징 계층에는 비디오의 프레임에 대한 시각적 특징 정보와 프레임에 나타난 객체들에 대한 시각적 특징 정보를 표현하였다. 이를 통해 프레임들 사이에 유사도 검색과 객체들 사이의 유사도 질의가 가능하다. 또한 객체에 대한 공간적인 위치 정보에 대한 거리 관계를 이용하여 유사도 질의를 수행할 수 있다.

(4) 시공간 관계에 기반한 내용 질의

제안하는 모델링의 시각적 특징 계층에는 동일한 프레임에 나타난 객체들에 대한 공간적인 위치가 표현되어 있다. 또한 의미적 특징 계층에는 의미적 내용들에 대한 시간 간격과 공간적 위치가 표현되어 있다. 이를 이용하여 시공간 관계를 통한 내용 질의를 수행할 수 있다.

5. 결론 및 향후 연구

본 논문에서는 비디오에 대한 효과적인 접근을 지원하기 위한 새로운 비디오 모델링 기법을 제안하였다. 또한 객체와 사건 또는 구조적 계층 간의 시공간적인 관계를 정의하였다. 이를 통해 다양한 형태의 질의를 지원할 수 있도록 하였다. 향후 연구 방향으로 시공간 관계를 통한 비디오 검색을 효과적으로 수행하기 위한 인덱싱 방법에 대해 연구를 진행할 것이다.

참고문헌

[1] Y. M. Park, Y. K. Kim, S. I. Jin, W. Choi, "Hierarchical Structure-based Metadata Model for Video Database Applications," ISCA COMPUTERS AND THEIR APPLICATIONS, pp.242-245, 1998
 [2] 최지희, 용환승, "시간 관계성을 기반으로 한 비디오 데이터 모델의 설계 및 구현", 한국멀티미디어학회 논문지, 제2권 제3호, pp. 252-264, 1999
 [3] 윤미희, 윤용익, 김교정, "비디오의 의미검색과 유사성검색을 위한 통합비디오정보시스템", 정보처리논문지, 제6권 제8호, pp.2031-2041, 1999
 [4] J. F. Allen., "Maintaining Knowledge about Temporal Intervals", Comm. ACM, 26(11), pp.832-843, 11, 1983
 [5] J. Z. Li, M.T. Ozsu, and D. Szafron, "Modeling of Moving Objects in a Video Database", Pro. IEEE International Conference on Multimedia Computing and Systems, pp.336-343, 1997