비디오의 객체 움직임 이해를 위한 시공간 관계 표현

최준호[0], 조미영, 김판구
조선대학교 대학원 컴퓨터학과
{spica[0], irune80, pkkim}@chosun.ac.kr

Representation of Spatio-Temporal Relations for Understanding Object Motion in Video

Junho Choi[0], Miyoung Cho, Pankoo Kim
Dept. of Computer Science, Chosun University

요 약

비디오 데이터에서 의미적 인식을 위해 활용되는 요소 중 하나가 객체에 대한 움직임 정보로 이는 비디오 데이터에 대한 색인과 내용 기반 검색을 수행하는데 중요한 역할을 한다. 본 논문에서는 효율적인 객체기반 비디오 검색과 비디오의 움직임 해석을 위한 시공간 관계 표현 방법을 제시한다. 비디오의 객체 표현 방법은 Polygon-based Bounding Volume의 3차원 Mesh 모델을 생성한 후 이를 이용하여 비디오 내 개체의 구조적 내용을 저차원적 속성과 움직임에 대한 기본 구조로 활용하였다. 또한, 움직임 객체에 대해 시공간적 특성과 시각적 특성을 동시에 고려하여 표현되도록 하였다. 각 Vertex는 시각적 특징 중 일부분이고, 비디오 내 개체의 공간적 특성과 개체의 움직임은 Volume Trajectory로 모델링되고, 개체와 개체간의 시공간적 관계를 표현하기 위한 Operation을 정의한다.

1. 서론

최근 비디오 검색 관련 연구는 비디오 데이터로부터 의미적 특징을 자동으로 추출한 후 이를 이용한 내용 기반 검색이 가능하도록 하는 것이 주요 테마이다. 특히, 비디오 데이터에서 의미적 인식을 위해 활용되는 요소 중 하나가 객체에 대한 움직임 정보로 이는 비디오 데이터에 대한 색인과 내용 기반 검색을 수행하는데 중요한 역할을 한다. 본 논문에서는 효율적인 객체기반 비디오 검색과 비디오의 움직임 해석을 위한 메커니즘을 위한 논리적인 프레임워크를 제시한다. 비디오의 객체 표현 방법은 Polygon-based Bounding Volume이라는 3차원 Mesh 모델을 생성하여 이를 비디오 내 개체의 구조적 내용을 저차원적 속성과 움직임이다[1]. 본 논문에서는 시공간적 특성과 시각적 특성을 동시에 고려하여 표현되도록 하고자 한다. 각 Vertex는 시각적 특징 중 일부분이고, 비디오 내 개체의 공간적 특성과 개체의 움직임은 Volume Trajectory로 모델링되고, 개체와 개체간의 시공간적 관계를 표현하기 위한 Operation을 정의한다.

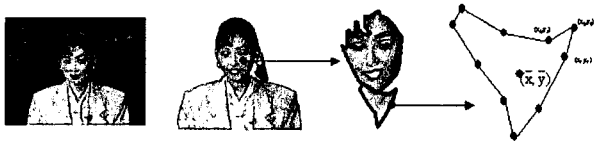
2. 비디오 내의 움직임 객체 영역 분리

2.1 객체 추출 초기화

비디오 내의 객체 추출을 위해서 먼저 임의로 지정한 윤곽선의 화소 밝기값과 색상정보를 이용하여 화소들의 분할과 병합과정에 의해 군집화를 수행하고 워터쉐드 기법 등을 적용하여 정확한 객체의 경계를 추출한다. 이전 프레임에 포함된 객체의 이동경로를 파악하기 위해 동영상 객체에 대한 움직임 벡터를 구하고, 이 벡터로부터 3차원 병진, 회전, 선형변환을 대응하는 움직임 파라미터를 산출하는 과정과 압축된 비디오 데이터를 복원하고 움직임 벡터를 추출한다.

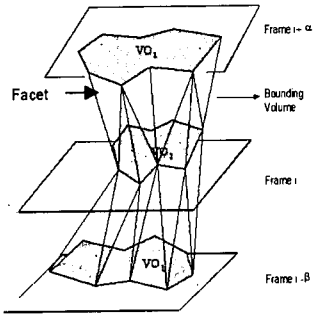
2.2 객체의 Polygon Mesh 화

비디오 시퀀스로부터 Shot 정보가 성공적으로 추출이 되고 나면, 다음 단계로서 생각할 수 있는 것이 객체기반 비디오 분석 단계이다. 비디오 객체를 하나의 프레임으로부터 지정할 수가 있다. 그러나 자동적인 의미 해석 기술이 아직 완전하지 않으므로, 여기에서는 시각적 객체에 대한 의미해석을 사용자가 제공하여 텍스트 정보로서 유지하고 있다고 가정한다. 프레임에 나타나는 물리적 객체의 공간적 속성은 Bounding Box나 Bounding Region으로서 추출할 수가 있다. 여기에서 프레임 속의 비디오 객체에 대한 Bounding Region은 bounding volume으로 확장된다. 또한, 비디오 데이터에 존재하는 개체들의 수가 방대하기 때문에 저장 공간을 줄이면서 처리 시간을 빠르게 하기 위해 개체의 Polygon Mesh의 기하학적 구성 요소를 적게 가지면서 유사한 모양을 유지할 수 있는 Mesh를 간략하게 표현한다. 다음 (그림 1)은 비디오의 한 프레임에서 얼굴 영역 부분을 추출하여 외각 영역을 재구성한 예이다.



(그림 1) 개체의 윤곽선 추출 및 Polygon Mesh화된 개체

또한, Bounding Volume을 생성하기 위해서 Skinning 알고리즘을 사용하였는데, Skinning 알고리즘은 경계 데이터를 Polygon(3차원 볼륨)으로 전환한다. (그림 2)를 살펴보면, 각각의 면은 Bounding Volume에서 패싯(Facet)이라 불리는 삼각형이다. 따라서 각 간격에서의 시각적 객체는 경계 영역과 패싯의 집합이다.



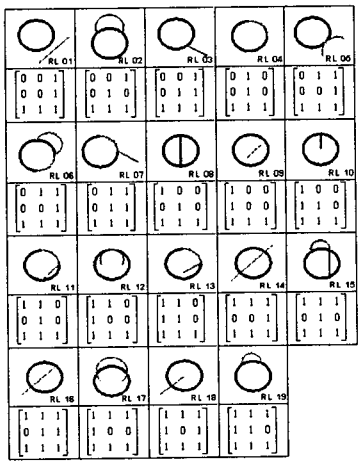
(그림 2) Bounding Volume

3. 객체간의 움직임 관계 정의

3.1 객체간의 시공간 관계 정의

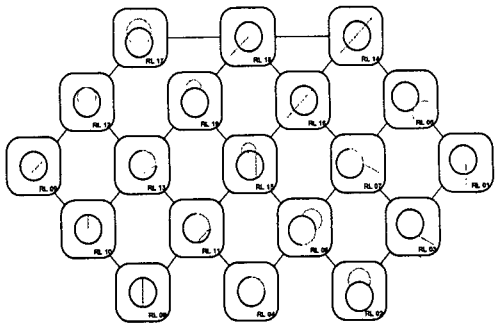
디지털 영상 내 객체간의 시공간적 관계의 정의는 객체의 의미적 인덱싱(Semantic Indexing), 검색, 의미적 객체 부호화를 통한 객체단위의 비디오 조작, 편집, 의미적 영상 합성 등을 위하여 필요한 기술이며, 영상 내의 객체들에 대한 움직임을 추정하기 위하여 큰 영상을 단위 영상으로 분할하여, 단위 영상 내 객체들에 대한 움직임(motion trajectory)과 객체간의 관계 등을 이용하여 Spatial-temporal 인덱싱을 수행한다[2].

기존의 비디오 객체간의 위치 관계 표현은 영역과 영역간의 관계만을 다루고 있지만, 본 논문에서는 동적인 움직임 객체를 점으로, 정적인 객체를 영역으로 선과 영역간의 위치 관계로 설정하였다.



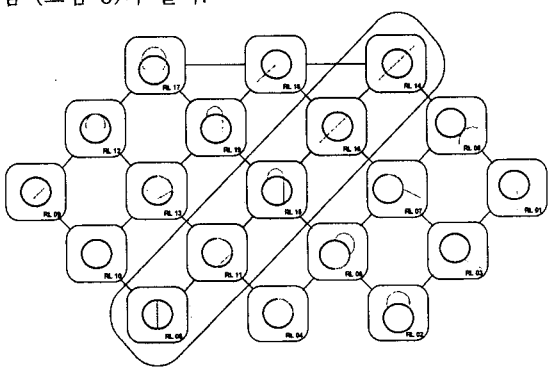
(그림 3) 19가지 선-영역 위상 관계

(그림 3)에서 정의한 선-영역 위치 관계의 선의 경계와 내부를 Push와 Pull을 반복하는 과정을 통해 행렬 각각의 값의 차이를 비교하여 위상 근접 그래프로 나타내면 (그림 4)와 같다.



(그림 4) 위상 근접 그래프

이는 움직임에 대한 정확한 위상 관계의 표현으로 의미 및 내용 기반 비디오 검색에 활용될 수 있을 것이다. 예를 들어, 객체가 화면을 가로지르는 움직임 'cross'에 대한 표현을 위해 위치 관계를 근접 그래프에 표현하면 다음 (그림 5)와 같다.



(그림 5) 'Cross'를 표현하기 위한 위치 관계

이와 같은 비디오 내의 객체간의 움직임 관계를 표현하기 위해 본 논문에서는 다음 (표 1)과 같이 정의하였다.

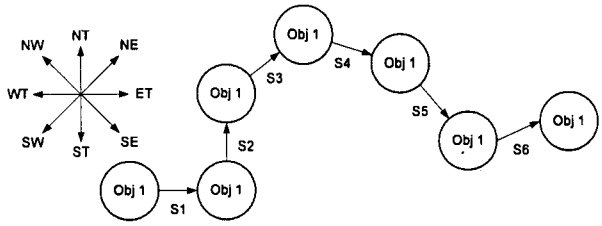
(표 1) Spatio-Temporal 관계 정의

Temporal-Interval Relations	{before, meet, overlap, during, start, finish, equal}
Strict Directional Relation	{north, south, west, east}
Mixed Directional Relations	{northeast, southeast, northwest, southwest}
Positional Relations	{left, right, above, below}
Topological Relations	{inside, outside, overlap}

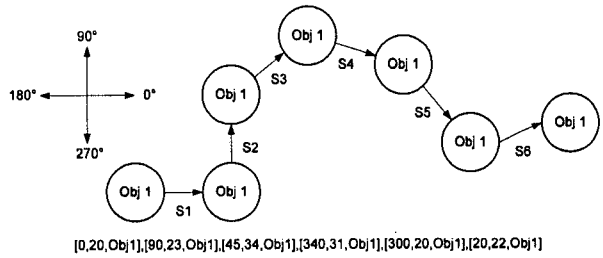
3.2 객체간의 방향, 속도 관계 모델링

비디오 데이터 내의 분할된 개체를 기반으로 시간 축상의 영역의 흐름을 찾을 수가 있다. 이를 위해서 먼저 기준 프레임의 각각의 개체에 대해 다음 프레임의 영역들에 대한 대응관계(correspondence)를 찾아야 하는데, 이를 위해서 영역의 밝기나 사이즈 혹은 비슷한 위치에 있는 영역과의 차이를 나타내는 임계치를 설정하도록 한

다. 이러한 개체간의 방향 및 속도 등을 파악하는 것은 그 개체와 주위 개체와의 관계성 의미 파악에 중요한 데이터로 사용될 수 있다.



(그림 6) 움직임 객체의 방향 관계 모델링



(그림 7) 움직임 객체의 움직임 궤도 모델링

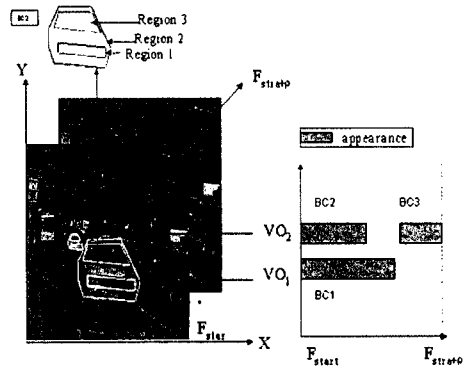
움직임 객체의 방향 관계 모델링은 (그림 6)과 같이 단일 움직임 객체의 궤적을 움직임 거리와 움직임 방향으로 나타내고, 움직임 객체의 움직임 궤도 모델링은 (그림 7)과 같이 단일 움직임 객체의 궤도를 움직임 거리와 움직임 방향으로 나타내지만 방향 관계 모델링 방법과 달리 거리는 실제 거리를 백분율로 표시한 상대적 거리, 움직임의 방향 관계는 각도(Angle)를 사용하여 표현하고 있다. 움직임 객체 모델링은 객체를 중심으로 설계하고, 객체의 움직임을 MBR로 표현하여, 객체 사이의 상대적 거리 유사를 측정할 수 있다.

3.3 비디오 데이터의 장면 데이터 모델링

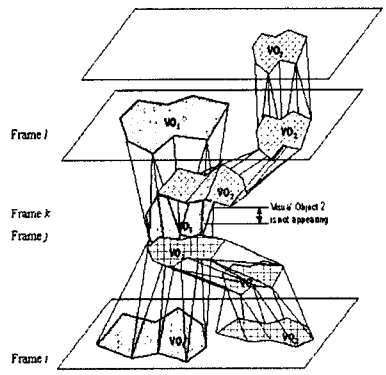
정확한 간격의 시각적 객체의 정의로부터 Shot의 형식적 데이터 모델을 정의할 수 있다. 정확한 간격의 Shot S는 다음과 같이 정의된다.

$$S=\langle S_{id}, VO, BV, SR, SD\rangle$$

Sid은 Shot id이다. VO는 전에 정의된 비디오 객체의 집합, 그리고 BV는 Bounding Volume의 집합이다. SDs는 Shot의 의미적 요약이고, SR은 Shot에서 VO와 BV의 Spatio-Temporal 관계의 집합이다. Shot 데이터 모델은 비디오 데이터의 시간과 공간적인 면을 포함한다. 이 Shot 데이터 모델은 클립 안에서 이벤트를 서술하거나 사용자들에게 복잡한 견해를 구성할 수 있도록 하는 것이 중요하다. 예를 들어, 각 비디오 객체에 대해 우리는 형태 기반의 Bounding Volume을 생성한다. (그림 8)에서 첫 번째 시각적 객체 "노란 차"는 주어진 Shot에서 일률적인 연속적인 모양을 가지고 있다. 그러나 두 번째 비디오 객체 "작은 노란 차"는 첫 번째 객체가 순간 두 번째 객체를 숨기기 때문에 Shot에서 2가지 다른 모양을 가지고 있다. 이 Shot에서, 두 비디오 객체를 위한 3개의 Bounding Volumes을 가진다. 그러나 Bounding Volumes BC2와 BC3는 의미적으로 동등한 비디오 객체로부터 생성된다.



(그림 8) 비디오 장면의 객체 추출 및 위치 관계



(그림 9) Bounding Volume 생성

4. 결론

본 논문에서는 비디오 내의 개체 움직임을 파악하여 그 내용을 3차원 Mesh 모델을 생성하고, 그 개체간의 시공간적 관계 표현을 위한 Polygon-based Bounding Volume(2차원, 3차원 Mesh)을 생성한 후 이를 활용하여 개체나 개체간의 움직임을 인식할 수 있는 프레임워크를 제안하였다. 이를 위해 윤곽 데이터를 다수의 다각형으로 변환하는 Skinning 알고리즘을 사용하였고, 볼륨 생성을 위해 Triangularization 기법을 사용하였다. 이를 기반으로 하여 각각의 동영상 내의 객체에 대한 기본적인 움직임을 이해하기 위한 시공간 관계를 정의하였다.

참고 문헌

[1] M. R. Naphade, I. V. Kozintsev, and T. S. Huang, "A factor graph framework for semantic video indexing," IEEE Trans. Circuits Syst. Video Technol., vol. 12, pp. 4052, Jan. 2002.

[2] Dierba, C.: Content-based multimedia indexing and retrieval. IEEE Multimedia9 :2002.

[3] M. Andrea Rodriguez, Max J. Egenhofer, Andress D. Blaser, "Query Pre-processing of topological Constaints:Comparing a Composition-based with Neighborhood-Based Approach", SSTD 2003, LNCS 2750, pp. 362-379, 2003.