

주석 및 내용 기반 뉴스 동영상 검색을 위한 데이터베이스 스키마 및 질의 유형 설계

○
전미경*, 김인홍*, 강현석*
*경상대학교 컴퓨터과학과/전산 개발 연구소

Design of Database Schema and Query Type for Supporting Caption- and Content-based News Video Searches

○
Mi-Gyung Jun*, In-Hong Kim*, Hyun-Syug Kang*
*Dept. of Computer Science/The Institute of Computer Research and Development,
Gyeong-Sang National University

요 약

일반적으로 동영상 검색을 위해 주석 기반과 내용 기반 검색을 사용하는데, 주석 기반 검색은 사용자의 주관이 개입되어 일관성을 잃기 쉽고, 내용 기반 검색은 동영상 데이터가 담고 있는 의미가 추출되기 어렵다는 단점을 가지고 있다. 그래서, 본 논문에서는 이 두 검색 기법을 상호 보완하여 검색의 효율성과 정확성을 높이기 위해 통합 동영상 데이터 모델(IVDM)을 제안하고, 이것을 기반으로 뉴스 동영상 검색을 위한 데이터베이스 스키마와 질의 유형을 설계한다. 이 모델은 동영상 데이터를 계층적으로 구조화한 형태로 상위 수준에서는 주제별로 부여된 메타 정보로 주석 기반 검색을 지원하고, 하위 수준에서는 동영상 데이터에서 색깔, 모양, 움직임, 질감 등의 특징 데이터를 추출하여 내용 기반 검색을 지원한다.

1. 소개

동영상 검색을 위한 연구들은 주석 기반 검색[2, 3, 11]과 내용 기반 검색[4, 6, 9]으로 나누어진다. 내용 기반 검색은 동영상 데이터에서 특징 데이터 즉 모양, 색깔, 질감, 움직임 등을 자동 추출하여 이를 바탕으로 검색하는 것이다. 따라서 사용자의 주관이 개입되지 않아 객관적인 검색이 가능하고, 특정 정의역에 대해서는 좋은 검색 결과를 나타낸다. 그러나 동영상 데이터가 담고 있는 일반적인 의미를 추출하기 어렵다는 단점이 있다. 반면 주석 기반 검색은 동영상 자료가 담고 있는 의미를 사람이 먼저 파악한 후 이를 자연어를 이용하여 표현하고, 이를 바탕으로 검색하는 방법이다. 이 방법은 자동화된 방법으로 알아내기 힘든 동영상 자료의 여러 가지

의미를 쉽게 모델링할 수 있게 하고 이를 검색에 이용할 수 있게 한다. 그러나 사용자의 관점에 따라 주석을 다르게 부여하거나 다르게 해석할 수 있으므로 동영상 자료에 대한 일관성을 잃기 쉽다. 특히 포괄적인 내용의 주석이 아닌 아주 세분화된 부분까지 주석을 부여하고자할 때는 더욱 더 일관성을 유지하기가 어려워진다. 따라서 이 두 접근법이 갖는 각자의 장점을 함께 수용하고 단점은 상호 보완하여 사용한다면 효과적일 것이다. 그리고 이러한 연구에서 동영상 정보를 압축된 상태에서 검색할 수 있어야 한다.

본 연구에서는 방대한 양의 동영상 데이터를 효율적인 방법으로 저장하고 검색할 수 있도록 일반적인 동영상 데이터를 분석하여 주석 기반 검색과 내용 기반 검색을 함께 사용할 수 있는 통합 동영상 데이

타 모델(IVDM)을 제시하고, 이 모델에 맞는 뉴스 동영상 데이터베이스 스키마 구조와 질의 유형 4가지를 설계한다.

논문의 구성은 다음과 같다. 2장에서는 동영상 검색 기법에 관련된 연구를 살펴보고, 3장에서는 주식 기반 검색과 내용 기반 검색을 동시에 가능하게 하는 통합 동영상 데이터 모델을 제안한다. 4장과 5장에서는 3장에서 제시한 모델에 기반하여 뉴스 동영상 검색을 위한 스키마를 설계하고, 이에 알맞은 질의 유형을 제시한다. 마지막으로 6장에서는 결론과 향후 연구 과제를 알아본다.

2. 관련 연구

동영상 데이터를 검색하는 방법에는 크게 내용을 기반으로 하는 검색과 주석을 기반으로 하는 검색으로 나눌 수 있다.

우선 동영상에 텍스트 형태의 정보(주석)를 부여하여 검색에 이용하는 주식 기반 검색 연구를 알아본다.

이 검색 기법의 대표적인 예로 CVIMS[3]와 VIRON[2]이 있다. 허진용 등이 개발한 CVIMS은 주석을 캡션과 그림 설명 정보로 나누어 관리하는 시스템이고, VIRON은 김기욱과 김형주가 단일의 주식 시스템 구조를 속성 스키마, 개념적 주식 구조, 동영상 인덱스 스트림 구조로 분할하여 동영상 데이터 질의 시스템을 모델링한 것이다. CVIMS와 VIRON은 가능한 객관적인 관점에서 주석을 부여하여 공유하도록 설계를 하였지만 여전히 주식 기반의 한계라 할 수 있는 사용자의 주관성을 완전히 배제하지는 못하고 있다.

이와 달리 사용자의 주관이 전혀 개입되지 않는 동영상 데이터의 움직임과 같은 시간 차원과 색깔, 모양, 질감과 같은 공간 차원의 특징을 추출해 내어 검색에 이용하는 내용 기반 검색에 관한 연구도 많이 이루어졌다.

먼저 내용 기반 검색 시스템에서 사용되는 질의 유형을 보면 키워드를 입력으로 사용하는 경우와 그림을 그릴 수 있는 툴(Tool)의 도움을 받아 사용자가 직접 스케치한 그림이나 화일로 존재하는 정지영상(화상)을 입력으로 사용하는 경우가 있다.

키워드를 사용한 예로 A. Ono[5] 등은 장면 묘사어(Scene Description Keyword)를 사용하여 170여개의 정치 화상을 검색하는 시스템을 제안하였고,

Y. Kim과 M. Shibata[13]는 동영상의 내용을 자연어를 인덱스 트리 형태로 표현하여 검색시 이용하였다. 반면 QBIC[6, 12]과 Chabot[10], VisualSEEK[7]은 화상의 종류 키워드, 키워드와 함께 텍스처, 스케치, 모양을 사용한 대규모의 데이터베이스 상에서 검색할 수 있도록 연구하였다.

그런데, 이러한 내용 기반 검색 시스템에 해당하는 QBIC, Chabot, VisualSEEK 등은 아주 제한적인 범위의 검색에는 효과적이지만 동영상의 특성상 방대한 양의 데이터가 담고 있는 일반적인 의미를 추출하지 못하기 때문에 현실적이지 못하다.

따라서 이러한 두 종류의 기법 즉, 주식 기반 검색과 내용 기반 검색을 통합적으로 사용할 수 있는 방안의 모색이 필요하다. 이때 사용자의 요구에 적합한 형태로 두 가지 검색 기법을 통합적으로 지원하기 위해서는 무엇보다 통합 데이터 모델의 개발이 선행되어야 한다.

최근, 이 두 기법들을 통합하는 형태의 검색을 위해 김기병 등[1]은 다중 계층 동영상 모델(Multi-Layered Video Model; MuVi)을 제안하였다. 그러나 이 모델은 동영상 데이터 자체와 부여된 메타 정보 및 추출된 특징 정보들이 혼합되어 일직선상의 계층을 이루고 있으며, 개념 계층인 주식 기반 검색을 위한 모델링이 미흡하다.

본 논문에서는 동영상의 물리적인 구조에서 의미 있는 개념 계층으로 분할하는 MuVi 모델과는 달리 동영상의 논리적인 데이터를 기준으로 사용자의 요구를 넓은 범위에서 좁은 범위로 줄여 갈 수 있도록 상위 수준에서는 여러 단계를 거쳐 주석을 부여하도록 분할하고(주식 기반 검색에 이용), 하위 수준에서는 동영상 데이터의 특징을 추출해 낼 수 있도록 분할하는 방법(내용 기반 검색에 이용)을 사용한다.

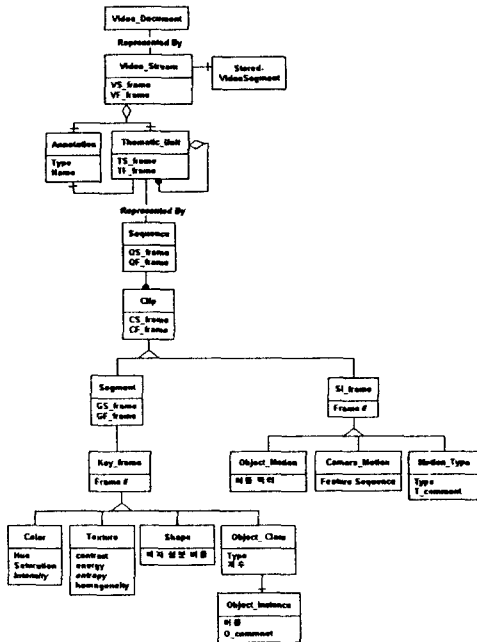
3. 통합 동영상 데이터 모델(IVDM)

동영상을 효율적으로 검색하고 관리하기 위해서는 압축된 동영상 자체, 그에 관련된 주석, 그리고 화상 분석 결과를 통합적으로 데이터베이스화하여 공유할 필요가 있다. 이를 위해서는 일반적인 표준 모델을 만들어 다양하고 방대한 양의 압축 동영상을 관리하는 작업이 요구된다. 따라서 이 장에서는 동영상 정보의 관리를 위한 통합 동영상 데이터 모델(Integrated Video Data Model, IVDM)을 제안한다.

이러한 IVDM의 의의는 동영상 데이터를 구조화

함으로써 다양한 동영상 데이터에 대해 자유로운 주석 기반 검색을 지원하고 동시에 내용 기반의 검색도 지원한다는 것이다. 그리고 이 두 가지의 단점을 서로 보완하고 장점을 극대화하기 위해 통합 형태의 검색도 동시에 지원한다. 이러한 작업으로 동영상 정보를 공유하여 사용할 수 있게 된다.

본 논문에서 제시하는 통합 동영상 데이터 모델 (IVDM)은 <그림 1>과 같이 구성된다. <그림 1>은 OMT 객체도[8]로 기술되었다.



<그림 1> 통합 동영상 데이터 모델(IVDM)

그림에서 Video_Document는 Video_Stream으로 다시 1:1 표현되며 이는 1개 이상의 Stored_VideoStream으로 구성되어 DB에 저장된다. 이 때 Video_Stream은 시작 프레임과 끝 프레임을 나타내는 두 가지 애트리뷰트를 가진다. Video_Stream과 Annotation 및 Thematic_Unit 사이는 부품 관계(part_of)로서, 1개 이상의 Annotation과 Thematic_Unit의 집합이 Video_Stream이 된다. Thematic_Unit은 다시 더 작은 Thematic_Unit을 포함할 수도 있고, 아닐 수도 있으며 Video_Stream과 마찬가지로 1개 이상의 Annotation으로 구성된다. 또 이 Thematic_Unit을 Sequence로 다시 표현하고, Sequence는 0개 이상의 Clip으로 이루어진다. 0개

이상이므로 Sequence 이하에 Clip이 없을 수도 있는데, 이것은 주석 기반의 검색만 가능하게 되는 것이고 적어도 하나의 Clip이라도 존재한다면 이는 내용 기반 검색까지 지원하게 되는 것이다. Clip과 Segment 및 SI_frame 사이는 일반화 관계(is_a)로서, Clip을 Segment나 SI_frame으로 다시 표현할 수 있다. SI_frame은 각각 Object_Motion, Camera_Motion 및 Motion_Type으로 표현할 수 있고, Segment는 Key_frame과 참조 관계(association)에 있다. 이 Key_frame은 다시 Color, Texture, Shape 및 Object_Type으로 표현할 수 있으며 마지막 Object_Type은 1개 이상의 Object_Instance로 구성된다.

4. 뉴스 동영상 데이터베이스 스키마 구조

이 장에서는 앞에서 제시한 IVDM을 이용하여 뉴스 동영상을 다루는 데이터베이스 스키마를 설계한다.

IVDM을 기반으로 한 뉴스 동영상 데이터베이스 스키마를 설계하기 위해 뉴스를 분석해보자. 하루에 방영되는 1회분 뉴스가 하나의 화일로 존재한다고 가정하면 그 뉴스는 정치, 사회, 경제, 스포츠 등의 주제로 분류해볼 수 있다. 다시 각 주제별로 여러 개의 사건들이 있고, 하나의 사건을 또 1명 이상의 기자들이 취재를 한다. 그리고, 한 기자가 취재한 내용도 장면을 기준으로 몇 번의 변화가 생긴다. 이러한 일반적 기준에서 설계된 데이터베이스 스키마는 <그림 2>와 같이 구성된다. 이 스키마 역시 OMT도의 형태로 기술되었다. 각 네모 상자의 윗 부분은 DB의 클래스명이고, 아랫 부분은 각 클래스의 속성들이다.

여기서는 여러 가지 동영상 데이터 중에서도 동영상의 대표격이라 할 수 있는 뉴스를 예로 구체적으로 설계하였지만 다른 종류의 동영상 데이터도 IVDM에 알맞게 video_kind 클래스 이하로 확장될 수 있다.

뉴스 동영상은 제일 상위에 뉴스(news) 클래스에 속성으로 뉴스 동영상의 시작 프레임과 마지막 프레임을 나타내는 s_frame과 f_frame이 정수형을 갖고, 제목(title), 방송날짜(broadcast_date), 방송시간(broadcast_time), 방송국(broadcast_station), 앵커

름(anchorman, anchorwoman), 톱뉴스(topnews)와 같은 뉴스에 관한 전반적인 정보들과 축소그림(thumbnail_frame)이 속성으로 존재한다. 그 아래의 주제(theme_unit) 클래스는 news 클래스의 속성들을 상속받는다. theme_unit 클래스의 분류(t_kind) 속성에는 경제, 사회, 스포츠, 정치 등의 문자열 값이 들어가고, 분류된 주제별 동영상의 범위를 나타내는 시작 프레임과 끝 프레임인 ss_frame과 sf_frame을 속성으로 갖는다. 이 외에 주제별 톱 사건(t_topnews), 사건 개수(t_event_num)도 속성이 된다. <그림 2>에서 보는 바와 같이 theme_unit 클래스의 속성들을 상속받는 사건(event_unit) 클래스가 있고, event_unit 클래스의 속성들을 상속받는 기자(reporter_unit) 클래스가 있다. event_unit과 reporter_unit 클래스 모두 시작 프레임과 끝 프레임을 나타내는 속성과 제목, 사건에 해당하는 누가(who), 언제(when), 어디서(when), 무엇(what), 어떻게(how), 왜(why)가 속성으로 포함된다. reporter_unit 클래스에는 기자의 이름도 속성으로

더해진다.

news 클래스부터 reporter_unit 클래스까지가 주석 기반 검색을 지원한다. 그 아래로 scene_unit, segment, key_frame 클래스 순서와 scene_unit, si_frame 순서로 상속을 받는다. 최종적으로 color, shape, texture, object_type은 같은 레벨로 이들 위의 모든 클래스들을 상속받는다. 마찬가지로 camera_motion, object_motion, motion_type도 단말 클래스로 이들 위의 모든 클래스들을 상속받는다. key_frame 클래스와 si_frame 클래스의 상속받는 단말 노드의 속성 값은 자동화된 방법으로 동영상을 분석한 결과 나온 데이터 수치들이다.

5. 질의 유형

이 장에서는 앞에서 제시한 스키마 구조에 적합한 질의 유형과 질의 처리 과정을 보여준다.

5.1 질의의 종류

질의 유형은 핵심 클래스인 theme_unit, key_frame, si_frame을 기반으로 4가지로 분류된다.

1) Type-1 : Thematic_Unit 기반의 질의(주석 차원)

큰 주제에서 작은 주제로 단계별로 부여된 주석을 기반으로 검색한다.

예1) "1998년 7월" "경기 지역" "수해" 상황을 검색하라

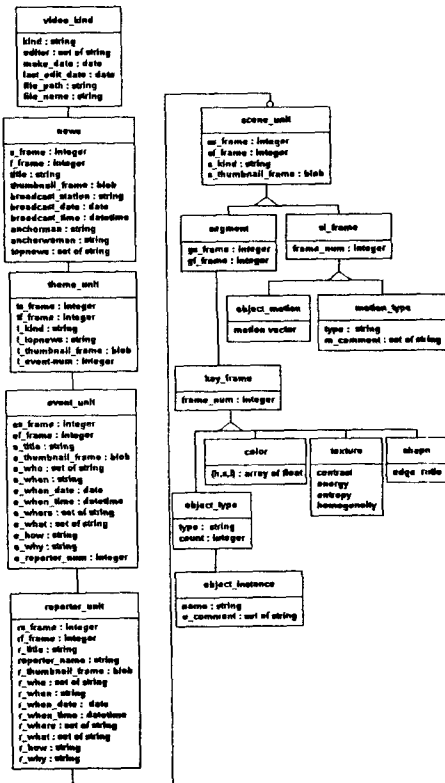
```
SELECT es_frame, ef_frame FROM event_unit
WHERE (e_when >= '1998/07/01' AND e_when <
'1998/08/01') AND (e_where IN '경기도
') AND e_what LIKE '수해%';
```

예2) "박세리" 선수의 "우승"에 관한 기사를 검색하라

```
SELECT es_frame, ef_frame FROM event_unit
WHERE (e_who IN '박세리') AND (e_what LIKE
'우승');
```

2) Type-2 : SI_frame 기반의 질의(시간 차원)

시간의 흐름에 따른 객체의 움직임에 대해 검색한다. 검색하고자 하는 객체의 움직임 형태를 키워



<그림 2> 뉴스 동영상 데이터베이스 스키마

드로 질의할 수도 있고, 실제 그 장면이 나타나는 동영상의 질의로 줄 수도 있다.

예3) “뛰는” 모습이 담긴 동영상을 검색하라
(Keyword 기반).

```
SELECT cs_frame, cf_frame FROM motion_type
WHERE type LIKE '뛰다';
```

예4) 뛰는 모습이 담긴 동영상을 검색하라(분석 알고리즘을 이용).

```
SELECT cs_frame, cf_frame FROM object_motion
WHERE !QUERY: 동영상 이동벡터 - DB: 동영상 이동
벡터; < 임계치;
```

3) Type-3 : Key_frame 기반의 질의(공간 차원)

특정 객체가 나타나는 장면을 검색한다. 화상에 나타나는 객체의 타입과 이름, 색깔, 모양, 질감 등을 키워드로 질의할 수도 있고, 직접 그림을 그리는 틀을 이용하여 스케치한 화상이나 파일의 형태로 존재하는 화상을 가져와서 검색할 수도 있다.

예5) “신창원”의 몽타주가 나타나는 장면 검색을 검색하라(Keyword 기반).

```
SELECT gs_frame, gf_frame FROM object_instance
WHERE name = '신창원' AND ('몽타주' IN
O_commet);
```

예6) 신창원의 사진 화상으로 검색으로 검색하라
(분석 알고리즘 이용).

```
SELECT gs_frame, gf_frame FROM color,
texture, shape
WHERE !QUERY: 화상의 특징값 - DB: 화상의 특징값
! < 임계치;
```

4) Type-4 : 통합된 질의

앞의 3가지 타입을 통합한 형태의 질의이다. 이 통합 형태는 하나의 타입형 질의로 검색한 결과물 도메인으로 다른 타입의 검색을 하므로 단계별로 검색 범위를 줄여 주게되어 더 빠르고 정확한 검색이 가능하다.

예7) “1998년 8월”에 “박찬호” 선수가 “던지는” 모습이 담긴 뉴스를 검색하라(Type-1과 Type-2의 결합).

```
SELECT es_frame, ef_frame FROM motion_type
WHERE (e_when_date>='8/1/1998') AND
```

```
(e_when_date<'9/1/1998') AND
(e_who IN '박찬호') AND type = '던지다';
예8) “박세리” 그리고 골프를 치는 사람이 있는
화상으로 검색하라(Type-1과 Type-3의 결
합).
```

```
SELECT es_frame, ef_frame FROM color,
texture, shape
WHERE (e_who IN '박세리') AND (!QUERY: 화상의
특징값 - DB: 화상의 특징값! < 임계치);
```

5.2 질의 처리 알고리즘

사용자의 질의를 처리하는 과정을 개괄적인 알고리즘으로 나타내면 다음과 같다.

[알고리즘 1] 사용자 질의 처리 알고리즘

CASE Type OF {질의의 유형에 따라 4가지로 분개}

Type-1 : {질의 유형이 주석기반 질의이면,}

Result = Search(Q); {단순 SQL문으로 검색}

Type-2 : {질의 유형이 SI_frame 기반의 질의이고,}

IF Keyword 방식 {키워드형 질의이면,}

THEN Result = Search(Q); {SQL문을 통해 DB의 자료를 검색}

ELSE BEGIN{키워드 방식이 아니면,}

Result = Analysis(Q_Video); {질의 동영상 분석}

Result = Search(Result); {결과를 DB에 있는 내용과 비교 검색}

END

Type-3 : {질의 유형이 Key_frame 기반 질의이고,}

IF Keyword 방식{키워드형 질의이면,}

THEN Result = Search(Q); {SQL문을 통해 DB의 자료를 검색}

ELSE BEGIN{키워드 방식이 아니면}

Result = Analysis(Q_Image); {질의 화상 분석}

Result = Search(Result); {분석한 결과물 DB에 있는 내용과 비교 검색}

END

Type-4 : {질의 유형이 혼합된 형태이면}

BEGIN

Result = Search_Type1(Q); {주석기반의 검색이

```

        있으면 먼저 검색}
Result = Search_Type2(Q) UNDER Result:
{그 검색 결과의 범위 내에서 Type-2 형태의
  검색}
Result = Search_Type3(Q) UNDER Result
{다시 이 검색 결과의 범위 내에서 Type-3 형
  태의 검색}
END
OTHERWISE{위의 4가지 유형의 질의가 아니면}
  Error{에러 처리}
END.
    
```

6. 결론 및 향후 연구 과제

최근 방대한 양의 멀티미디어 정보가 쏟아져 나오
고 있는데, 그 중에서도 동영상 정보가 큰 비중을
차지한다. 따라서 동영상 정보에 대한 검색 요구가
많아지고 있다.

본 논문에서는 이러한 동영상 정보를 효율적으로
검색하기 위해, 주석 기반 검색과 내용 기반 검색을
모두 수용하는 통합 동영상 데이터 모델(IVDM)을
제시하였다. 그리고, 이 모델에 기반한 뉴스 동영상
데이터베이스 스키마를 설계하고 이에 알맞는 질의
유형 4가지를 제안하였다.

IVDM은 동영상의 논리적인 데이터를 기준으로
사용자의 요구를 넓은 범위에서 좁은 범위로 줄여
갈 수 있도록 상위 수준에서는 여러 단계를 거쳐 주
석을 부여하도록 분할하고(주석 기반 검색에 이용),
하위 수준에서는 동영상 데이터의 특징을 추출해 낼
수 있도록 분할하는 방법(내용 기반 검색에 이용)을
사용한다. 우리는 IVDM에 기반하여 뉴스 동영상 데
이터베이스 스키마와 질의 유형을 설계해봄으로써
그 적합성을 입증하였다.

한편 본 논문에서는 인덱스 정보만 데이터베이스
화해서 관리 가능하도록 설계되었다. 차후에는 동영
상 데이터 자체도 DB에 구조화하여 저장할 수 있도
록 확장 설계되어야 한다.

참고문헌

[1] 김기병, 김형주, "내용 기반 검색 및 주석 기반 검색을
통합하는 비디오 데이터 모델의 설계 및 구현," 정보
과학회 논문지(C), 제3권 제2호, pp. 115-126, Apr.
1997.

[2] 김기욱, 김형주, "비디오 주석 시스템의 설계 및 구
현," 정보과학회 논문지, 제24권 제6호, pp. 588-597,
June 1997.

[3] 허진용, 김인홍, 배종민, 강현석, "MPEG-2 압축 동영
상 정보 관리 시스템의 설계 및 구현," 정보처리학회
논문지, 제5권, 제6호, pp. 1431-1440, June 1998.

[4] A. D. Bimbo and P. Pala, "Visual Image Retrieval
by Elastic Matching of User Sketches," IEEE Tran.
PAMI, Vol. 19, No. 2, pp. 121-132, Feb. 1997.

[5] A. Ono, M. Amano, M. Hakaridani, T. Satou, and
M. Sakauchi, "A Flexible Content-based Image
Retrieval System with Combined Scene
Description Keyword," IEEE, pp. 201-208, June
1996.

[6] I. A. R. Center, "Query by Image and Video
Content: The QBIC System," IEEE Computer, 28(9),
pp. 23-32, Sept. 1995.

[7] J. R. Smith and S. Chang, "VisaulSEEK: a Fully
Automated Content-based Image Query System,"
ACM Multimedia '96, Nov. 1996.

[8] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, and
W. Lorensen, "Object-Oriented Modeling and
Design," Prentice-Hall, 1991.

[9] R. K. Srihari, "Automatic Indexing and
Content-based Retrieval of Captioned Images,"
IEEE Computer, pp. 49-56, Sept. 1995.

[10] V. E. Ogle and M. Stonebraker, "Chabot: Retrieval
From a Relational Database of Images," IEEE
Computer, 28(9), pp. 40-48, 1995.

[11] W. E. Mackay, "EVA: An Experimental Video
Annotator for Symbolic Analysis of Video Data,"
SIGCHI Bulletin, 21: 68-71, Oct. 1989.

[12] W. Niblack, R. Barber, W. Flickner, E. Glasman, D.
Petkovic, C. Faloutsos, and G. Taubin, "The QBIC
Project: Querying Image by Content Using Color,
Texture, and Shape," Proc. SPIE Storage and
Retrieval for Image and Video Database, pp.
173-187, Feb. 1993.

[13] Y. Kim and M. Shibata, "Content-Based Video
Indexing and Retrieval - A Natural Language
Approach," IEICE Trans. Inf. & Syst., Vol.
E79-D, No. 6, pp. 695-705, June 1996.