

비디오의 의미검색과 유사성검색을 위한 통합비디오정보시스템

윤 미 희[†] · 윤 용 익^{††} · 김 교 정^{†††}

요 약

본 논문에서는 비정형, 대용량의 비디오데이터의 특징기반 검색과 주석기반 검색을 통합하여 다양한 사용자의 의미검색을 지원하고, 유사성 질의를 지원하는 통합비디오정보시스템(Hybrid Video Information System : HVIS)을 제안한다. HVIS는 메타데이터 모델링을 위해 한편의 비디오를 비디오 다큐먼트, 시퀀스, 장면, 객체로 나누고 물리적인 비디오스트림을 위한 원시데이터계층(raw_data layer)과 주석기반 검색, 특징기반 검색, 유사성 검색을 지원하기 위한 메타데이터계층(meta_data layer)의 두개의 계층을 가진 통합 객체지향 메타데이터모델(Two layered Hybrid Object-oriented Metadata Model : THOMM)과 이 모델을 기반으로 주석기반 질의, 특징기반 질의, 유사질의가 가능한 비디오질의언어(Video Query Language)와 질의를 처리하기 위한 비디오질의처리기(Video Query Processor : VQP)와 질의처리알고리즘을 제안한다. 특히 유사한 장면, 객체를 찾는 유사질의시 사용자의 관심을 고려한 유사성 정도를 나타내는 식을 제시한다. 제안된 시스템은 Visual C++, ActiveX와 ORACLE를 이용하여 구현되었다.

Hybrid Video Information System Supporting Content-based Retrieval and Similarity Retrieval

Mi-Hwi Youn[†] · Yong-Ik Yoon^{††} · Kio-Chung Kim^{†††}

ABSTRACT

In this paper, we present the HVIS(Hybrid Video Information System) which bolsters up meaning retrieval of all the various users by integrating feature-based retrieval and annotation-based retrieval of unformatted formed and massive video data. HVIS divides a set of video into video document, sequence, scene and object to model the metadata and suggests the Two layered Hybrid Object-oriented Metadata Model(THOMM) which is composed of raw-data layer for physical video stream, metadata layer to support annotation-based retrieval, content-based retrieval, and similarity retrieval. Grounded on this model, we presents the video query language which make the annotation-based query, content-based query and similar query possible and Video Query Processor to process the query and query processing algorithm. Specially, We present the similarity expression to appear degree of similarity which considers interesting of user. The proposed system is implemented with Visual C++, ActiveX and ORACLE.

† 준 회원 : 숙명여대 대학원 전산학과
 †† 종신회원 : 숙명여대 전산학과 교수
 ††† 정 회원 : 숙명여대 전산학과 교수
 논문접수 : 1999년 1월 21일, 심사완료 : 1999년 7월 28일

1. 서 론

최근 H/W 및 S/W의 발달로 인한 인터넷의 증가와 초고속정보통신망의 구축으로 인하여 Audio, Video와 같은 멀티미디어데이터를 위한 응용이 많이 등장하고 있다. 비디오데이터를 효율적으로 저장하고 처리하기 위해서는 비디오데이터가 가지고 있는 내용에 대한 정보를 데이터베이스에 저장하고, 사용자들의 다양한 질의를 처리할 수 있는 의미검색방법이 요구된다[1,2,3,4].

이런 비디오데이터베이스는 비디오 그 자체를 나타내는 Raw_data와 함께 의미검색을 위한 비디오데이터의 내용을 구조적으로 설명하는 내용의존적(content-dependent)인 메타데이터(meta-data)가 필요하다[5]. 메타데이터의 구성은 비디오가 가지고 있는 내용에 대한 정보를 컴퓨터에 의해 자동적으로 추출한 후 검색에 이용하는 특징기반검색기법(content-based retrieval) [1,5,6]과 비디오가 가진 의미를 사람이 먼저 파악한 후 자연어를 이용하여 표현하고 검색에 이용하는 주석기반검색(annotation-based retrieval)기법[2,9,10]이 있다. 특징추출에 의해 자동적으로 생성된 데이터에 대한 직접적인 질의를 수행하는 것은 거의 불가능하므로, 특징데이터를 사용자가 사용하기 위한 단어로 사상하여 메타데이터를 구성하고, 비디오가 가지고 있는 의미에 대해서는 사용자의 입력에 의해 메타데이터를 구성하여 검색에 활용한다면 효율적인 의미검색이 이루어질 수 있다.

비디오정보는 방대할 뿐만 아니라, 내포된 정보의 내용들이 다양하여 기존의 데이터베이스 검색방법으로는 사용자가 원하는 형태와 의미를 정확하게 처리하기 어려우므로, 비디오를 위한 새로운 검색방법이 필요하다. 다양한 사용자의 요구를 처리하기 위한 비디오 검색기를 위해서는 시공간질의[6], 색상질의, 키워드질의와 비디오 장면이나 객체의 유사 질의를 포함하는 비디오질의언어의 정의가 필요하며, 검색의 효율을 높이기 위한 질의완화(query relaxation)기법[7]이 필요하다. 특히 유사질의처리시 사용자가 원하는 가장 유사한 장면과 객체를 검색할 수 있어야하고, 또한 결과의 제시시 유사성의 정도를 확률로 표시할 수 있어야한다.

본 논문에서는 특징기반 검색과 주석기반 검색을 통합하여 다양한 사용자의 의미검색을 지원하고, 유사성 질의를 지원하는 통합비디오정보시스템(Hybrid Video Information System : HVIS)을 제안한다. 한편의 비디

오를 비디오 다큐먼트, 시퀀스, 장면, 객체로 나누고 물리적인 비디오스트림을 위한 원시데이터계층(raw-data layer)과 주석기반 검색, 특징기반 검색, 유사성 검색을 지원하기 위한 메타데이터계층(meta-data layer)으로 구성되며, 메타데이터계층은 다시 객체의 특징정보를 갖는 객체특징계층(object-feature layer)과 비디오에 포함된 의미를 표현하기 위한 개념계층(concept layer)으로 구성된 두개의 계층을 가진 통합 객체지향 메타데이터 모델과 이 모델을 기반으로 주석기반 질의, 특징기반 질의(특정질의), 유사질의가 가능한 비디오질의언어와 비디오질의처리기(Video Query Processor : VQP)와 질의처리알고리즘을 제안한다. 특히 유사질의 처리시 사용자의 관심을 고려하여 유사성의 정도를 확률로 제시하기 위한 유사성정도식(similarity degree expression)을 제안한다. 제안된 시스템은 Visual C++, ActiveX와 ORACLE를 이용하여 구현되었다.

본 논문의 구성은 2장에서는 관련연구로서 기존의 비디오데이터모델링 기법과 검색기법에 대해 고찰한다. 3장에서는 주석기반 검색과 특징기반 검색, 객체의 시공간 질의, 색상질의, 유사질의를 지원하는 두개의 계층을 가진 통합 객체지향 메타데이터모델(Two layered Hybrid Object-oriented Metadata Model : THOMM)을 제안한다. 4장에서는 제안된 THOMM을 기반으로 질의를 구성하기 위한 비디오질의언어와 유사성정도식을 제안하고, 5장에서는 제안된 비디오질의언어를 기반으로 생성된 질의어를 처리하기 위한 비디오질의처리기와 질의처리알고리즘을 제안한다. 6장에서는 구현 예를 보이고 7장에서는 결론을 맺는다.

2. 관련연구

VIDEO-STAR[2]는 비디오와 메타데이터의 공유와 재사용을 지원하기 위해 관계형 데이터 모델을 사용한 데이터베이스시스템이다. 메타데이터의 구성은 분할 기법을 통해 비디오데이터를 구조화하고, 구조화된 비디오데이터에 people, location, event의 메타데이터를 구성하고, 이를 다시 기본, 일차, 이차컨텍스트로 구분하여 재사용과 공유를 용이하게 하며, 고정된 속성을 통해 사용자가 쉽게 질의를 구성할 수 있다. 그러나 메타데이터가 people, location, event의 내용만으로 한정되어 있어 다양한 의미검색이 불가능하고, 질의를 처리하기 위해서는 비디오데이터베이스를 검색할 때 De-

compose, Map to composition 등의 대응연산을 수행하여 해당 비디오데이터와 메타데이터를 관련시켜야 하므로 질의처리가 복잡하다.

AVIS[8]는 비디오데이터의 내용을 쉽게 저장하고 검색할 수 있도록 개발된 비디오 데이터베이스시스템이다. AVIS는 비디오 내용에 나타나는 객체, 사건, 행위에 대한 메타데이터를 정의하고, 이를 비디오세그먼트트리에 연계시켜 비디오 내용에 대한 검색을 용이하게 하였으며, 새로운 데이터의 추가와 삭제에 대한 알고리즘을 제시했다. AVIS는 상위레벨의 내용을 중심으로 하는 주석기반 메타데이터의 유형만을 제공하며, 비디오에 대한 데이터 모델은 없으며, 하위 레벨의 내용을 나타내는 특징들에 대한 메타데이터가 없고, 명확한 질의언어를 제공하지 않는다.

OVID[9]는 비디오데이터를 모델링하고 검색하기 위해 객체지향접근방법을 객체지향 비디오 데이터베이스시스템이다. OVID는 의미 있는 장면(scene)을 비디오 객체(video object)라는 개념을 도입하고 각 비디오객체마다 포함된 내용을 설명하는 속성과 속성값을 부여하여 메타데이터를 구성하였다. OVID는 비디오객체들을 검색하기 위해 VIDEO SQL이라는 사용자 인터페이스를 통해 이루어진다. OVID는 비디오객체모델 스키마를 사용하지 않기 때문에 사용자는 각 비디오객체마다 완전한 속성과 속성값을 지정해야하고 비디오데이터의 구조가 일반화되어 있지 않으므로 질의형식을 지정할 수 없다. 그러므로 데이터베이스의 크기가 증가할수록 질의처리가 어렵다.

Venus[1]는 특징기반검색을 위해 개발된 비디오 데이터베이스시스템이다. 이미지처리 기술을 사용하여, 비디오의 각 프레임에 나타나는 객체들의 시간관계, 공간관계를 메타데이터로 구축하였다. Venus는 특징기반 질의언어인 CVQL(content-based video query language)을 제공한다. 객체에 대한 시공간 질의를 위해 비디오 함수를 사용한다. 그러나 비디오의 내용에 대한 상위레벨의 의미를 위한 메타데이터가 없으므로 다양한 질의를 구성할 수 없다.

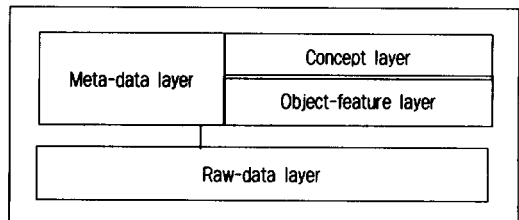
KMeD[6]는 의료분야에서 지식을 기반으로 이미지 검색을 위해 UCLA에서 개발된 프로토타입 비디오 데이터베이스시스템이다. KMeD는 관련된 이미지에 포함된 객체뿐만 아니라, 객체의 의미적인 기술을 표현하고 분류하기 위해 SEDM(spatial evolutionary data model), TEDM(temporal evolutionary data model)를

사용하여 모델링을 수행했다. 또한 사용자의 부정확한 질의에도 질의완화 과정을 통해 적절한 답을 제시할 수 있는 의료 이미지 데이터베이스시스템이다. 그러나 KMeD는 정확한 질의언어의 정의가 없으며, MRI와 CT와 같은 의료분야의 이미지데이터에 대해서만 적용 가능한 방법이다.

3. 비디오데이터를 위한 두개의 계층을 가진 통합 객체지향 메타데이터 모델

본 논문에서는 비디오정보를 다루기 위한 비디오 분할과정과 추출된 특징들을 색인하는 과정은 포함하지 않고 추출된 특징들과 비디오 내용에 대한 데이터의 모델링을 수행한다. 본 논문에서 제안하는 두개의 계층을 가진 통합 객체지향 메타데이터 모델(THOMM)은 원시데이터 계층(raw-data layer)과 비디오의 의미 검색을 위해 특징과 주석을 통합하는 메타데이터계층으로 구성되며 모델링을 위해 객체지향 모델링 기법을 사용하였다.

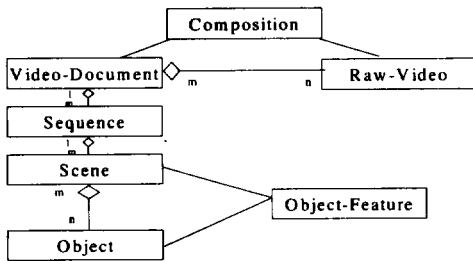
한편의 비디오다큐먼트는 필름이론(film theory)[10]에 의해 시퀀스(sequence), 장면(scene), 화면(shot), 프레임(frame)으로 구성된다. TV뉴스를 예로 들면 “1998년 10월26일 KBS 9시 뉴스”가 하나의 완전한 비디오다큐먼트가 되고, 하나의 기사가 시퀀스가 되며, 시퀀스는 앵커맨이 나오는 장면에서 다음 기사가 나오기 전까지의 장면의 집합으로 이루어진다[11]. 본 논문에서는 동일시간, 동일장소에서 일어나는 화면(shot)의 집합인 장면(scene)을 처리의 기본단위로 한다.



(그림 1) 두개의 계층을 가진 통합객체지향 메타데이터 모델

비디오를 위한 THOMM은 (그림 1)과 같다. 비디오의 물리적인 비디오스트림을 위한 원시데이터계층(raw-data layer)과 원시데이터의 내용을 기술하기 위한 메타데이터계층(meta-data layer)으로 구성되며, 메

타데이터계층은 주석기반 질의를 위한 개념계층(concept layer)과 특징기반질의(특징질의)를 위한 객체특징계층(object-feature layer)으로 구성된다. 또한 메타데이터계층을 이용하여 유사성질의를 지원할 수 있다. 원시데이터계층은 원시비디오객체로 구성되고 메타데이터계층은 비디오, 시퀀스, 장면, Object, 객체특징 객체로 구성된다.



(그림 2) THOMM의 객체들 사이의 관계

UML(Unified Modeling Language)를 사용한 THOMM의 객체들 사이의 관계는 (그림 2)와 같다. 원시비디오는 물리적인 프레임 스트림(frame stream)을 나타낸다. 비디오는 원시 비디오데이터를 공유할 수 있고 메타데이터도 재사용 될 수 있다.

THOMM에서 정의된 객체유형의 정의는 다음과 같다.

(1) 원시비디오 객체

디지털 형식으로 저장소에 저장되어 있는 실제 비디오 스트림을 표현하는 객체로써 비디오 형식, 상영시간의 물리적 속성으로 정의된다.

(2) 비디오 객체

하나의 제목을 가지는 한편의 비디오로 하나 이상의 시퀀스로 구성되며, 비디오를 구성하는 물리적 비디오 스트림인 원시비디오 객체와 "Composition" 관계를 갖는다.

(3) 합성 객체

비디오객체와 원시비디오 객체사이의 합성관계를 나타낸다. 비디오는 여러개의 원시비디오의 합성으로 구성되고, 하나의 원시비디오는 여러 비디오 객체에 의해 공유될 수 있다.

(4) Sequence 객체

비디오내에서 연결 가능한 이야기, 또는 하나의 완

전한 기사나 비디오에 나타난 의미 있는 사건을 나타낸다. Sequence는 비디오 구조객체의 속성들을 상속받고, 하나 이상의 Scene으로 구성된다.

(5) Scene 객체

동일시간, 동일장소에서 발생한 내용으로 구성된 데이터이다.

(6) Object 객체

Scene에 나타나는 관심 있는 Object를 표현한다.

(7) Object-feature 객체

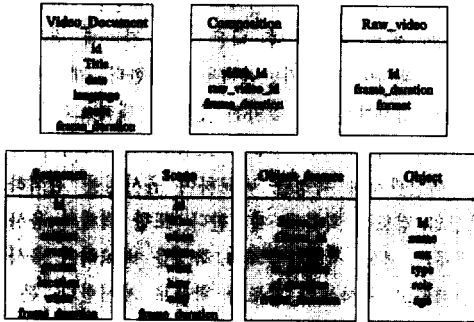
Scene과 Object 사이의 관계를 표현하기 위한 객체로 장면에 나타나는 관심 있는 객체의 색상, 시간적 관계, 공간적 관계를 표현한다. 이것은 영상처리를 통하여 자동으로 추출될 수 있다. 정밀도에 따라 다르지만 만일 하나의 이미지에서 하나의 특징을 추출하기 위해서 1초가 소요된다고 할 때, 질의 처리를 위한 시간은 데이터베이스에 1,000개의 이미지가 있으면 약 17분 이상, 10,000개의 이미지가 있으면 약 3시간 정도의 처리 시간이 요구된다는 것이다. 그러나 본 논문에서는 영상처리의 결과 없이도 사용자가 플레이되는 비디오를 보면서 관심 있는 객체의 위치를 최소 영역 사각형[12]으로 표시함으로써 해당하는 절대 좌표 정보 시스템에 의해 자동적으로 입력된다. 즉 절대 좌표 정보는 화면의 가장 왼쪽위의 좌표를(0,0)으로 오른쪽아래의 좌표를 (9,9)로 지정하고, 객체의 위치가 선택되면 왼쪽위의 좌표와 오른쪽아래의 좌표가 시스템에 의해 계산되어 입력된다. 또한 객체와 가장 유사한 색상을 색상테이블을 통해 선택함으로써 영상처리 없이도 공간적 관계와, 색상의 표현이 가능하다.

이러한 비디오 데이터를 기술하기 위한 속성을 다음과 같이 분류할 수 있다[12].

- ① 물리적 속성 : 프레임 듀레이션(frame duration) 등의 물리적인 속성.
- ② 내용독립적인 속성 : 비디오 제목, 제작사, 방송년월 일등과 같이 비디오 내용에 독립적인 속성.
- ③ 내용기반 속성 : 사건, 주제, 인물등과 같은 내용을 설명하는 속성으로 추상적인 의미를 기술하기 위한 속성과 객체의 색깔, 위치 등의 특징을 기술하기 위한 속성.

이러한 속성을 기반으로 각각의 클래스에 들어갈 속성들을 분류하면 <표 1>과 같다.

<표 1> 클래스의 속성



4. 비디오질의언어(video query language)

본 장에서는 통합비디오정보시스템에서 제안하는 비디오질의언어를 제시한다. 이 비디오질의언어는 THOMM에서 지원하는 메타데이터계층과 메타데이터계층의 클래스에 의해 구성된다.

4.1 비디오질의언어의 구문

HVIS에서 제공하는 비디오질의언어(video query language)는 메타데이터의 고정된 속성을 기반으로 한 주석기반 검색과 객체들간의 색상질의, 시공간질의, 특징기반질의, 유사질의를 제공하며, 구문은 다음과 같다.

```
FIND <result>
FROM <meta-db>
WHERE <predicate_clause>
```

FIND절은 사용자가 검색하고자 하는 대상으로 비디오다큐먼트, 시퀀스, 장면, 객체가 될 수 있다. FROM절은 질의의 탐색영역을 정의하며 WHERE절은 검색조건을 정의한다. 검색조건은 속성을 기반으로 한 조건절과, 객체들의 시공간조건절, 유사검색을 위한 유사관계식으로 정의된다.

비디오질의언어를 위한 BNF는 다음과 같다.

```
<result> ::= Video|Sequence|Scene|Object|Clip
<meta-db> ::= <video_clause>
<predicate_clause> ::= <predicate_stmt> | <logical_op> <predicate_stmt> |
<predicate_stmt> ::= <video_clause> | <sequence_clause> | <scene_clause> |
| <object_clause>
```

비디오다큐먼트에 대한 질의를 위한 <video_clause>는 다음과 같다.

```
<video_clause> ::= VIDEO <video_exp>
```

```
<video_exp> ::= <video_simple_exp> | <and> <video_simple_exp> |
<video_simple_exp> ::= <video_attribute> <relation_op> <value>
```

시퀀스에 대한 질의를 위한 <sequence clause>는 다음과 같이 정의된다.

```
<sequence_clause> ::= SEQUENCE <sequence_exp>
<sequence_exp> ::= <sequence_simple_exp> | <and> <sequence_simple_exp> |
<sequence_simple_exp> ::= <sequence_attribute> <relation_op> <value>
```

장면에 대한 질의를 위한 <scene_clause>는 다음과 같이 정의된다. <scene_clause>는 속성을 기반으로 한 검색과 주어진 장면과 유사한 장면을 검색하기 위한 유사성표현식으로 정의된다.

```
<scene_clause> ::= SCENE <scene_exp> | SCENE <scene_similar_exp>
<scene_exp> ::= <scene_simple_exp> | <and> <scene_simple_exp> |
<scene_simple_exp> ::= <scene_attribute> <relation_op> <value>
<scene_similar_exp> ::= SIMILAR_TO <scene_id>
```

객체에 대한 질의를 위한 <object_clause>는 다음과 같이 표현된다. <object_clause>는 공간질의와 시간질의를 표현할 수 있다.

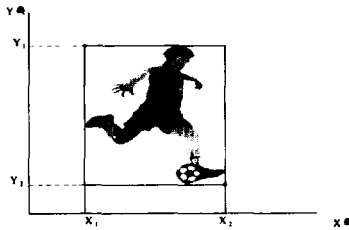
```
<object_clause> ::= OBJECT <object_exp> [at <position>] |
OBJECT <object_exp> <spatio_temporal> | OBJECT
SIMILAR_TO <object_id>
<spatio_temporal> ::= SPATIAL <spatial_relation> | TEMPORAL
<temporal_relation>
<position> ::= center|left|right|above|below|left_above|left_below|right_
above|right_below
<spatial_relation> ::= left|right|above|below|left_above|left_below|right_
above|right_below
<temporal_relation> ::= before|meets|overlaps|during|starts|finishes|equal
```

4.2 객체의 시공간관계

장면에 나타나는 객체의 공간관계와 시간관계를 표현하기 위해 메타데이터 계층의 객체특징 계층을 사용한다. 객체의 공간관계를 표현하기 위해 최소직사각형(Minimum Bounding Rectangular : MBR)방법을 사용한다.

본 논문에서는 영상처리의 결과로 얻어지는 객체의 공간관계를 표현하기 위해 사용자가 관심 있는 객체를 사각형으로 선택함에 의해 사각형의 왼쪽 위와 오른쪽 아래의 좌표가 X, Y의 좌표로 사상되어 구간을 형성하게 된다(그림 3). 객체의 절대공간 관계는 객체가 프레임에 나타나는 위치이다. X, Y축에 형성된 구간의 중심좌표를 이용하여 표현할 수 있다. 선각 X축을 중심으로 left, right를 정의하고, Y축을 중심으로 above와 below의 관계를 정의한다. 상대 공간관계는 한 장면에 나타나는 두 객체들간의 상대 공간 관계를 나타

낸다. 하나의 객체를 기준으로 다른 객체의 상대 공간 관계는 절대 공간관계와 같다.



(그림 3) 최소직사각형 방법

한 장면에서 나타나는 두 객체간의 시간 관계는 한 객체가 나타났다가 사라지는 시작 프레임번호와 종료 프레임번호와의 시간관계를 이용하여 표현 할 수 있다. Allen은 프레임구간으로 표현된 객체들간의 시간관계에 대한 표현과 추론을 위해 7가지의 시간관계를 제시하였다 [13]. 객체의 시간관계 표현을 위해 Allen이 제시한 7가지의 시간관계를 이용한다.

4.3 객체의 색상관계

비디오에 나타나는 객체의 색상은 평균색상과 대표색상으로 표현되며 색상히스토그램을 이용하여 구할 수 있다. 각각의 색상은 Red, Green, Blue의 3차원 벡터 (R,G,B)의 값으로 표현될 수 있다. 본 논문에서는 객체의 색상 입력시 그 객체가 가지는 대표색상을 색상테이블에서 선택하면 선택된 색상의 RGB값이 객체 특징계층의 색상의 속성값으로 입력된다.

색상질의를 객체특징계층을 사용하여 지원되는 유사질의로 사용자가 색상테이블에서 원하는 색상을 선택할 수도 있고, 색상값을 직접 입력할 수도 있다. 선택된 색상이 RGB값으로 사상된 후 색상을 비교한다. 색상의 비교식은 식 (1)과 같다. 계산된 Sim_{col} 값이 임계값(30)이하이면 유사한 색상으로 간주하고 결과를 제시한다. 이 방법은 색상의 유사성을 표현하는 가장 일반적이고 간단한 방법이다. 그러나 색상의 유사성을 위한 어떤 식도 사용 가능하다.

$$Sim_{col} = |R_q - R| + |G_q - G| + |B_q - B| \quad (1)$$

(단 R_q, G_q, B_q 는 질의로 입력된 RGB값, R, G, B 는 검색대상 객체의 RGB값)

4.4 유사관계

유사질의는 주어진 장면과 가장 유사한 장면이나 주

어진 객체와 가장 유사한 객체를 검색하기 위해 매타데이터계층을 사용한다. 첫번째로 개념계층을 사용하여 주석기반검색을 수행한 후 나온 검색 결과에 객체특징계층을 이용하여 특징기반검색을 수행한다. 개념계층의 주석기반 검색에 의해 찾아진 장면에 대해 그 장면에 포함된 객체와 주어진 장면에 포함된 객체사이의 공간관계, 색상관계의 특징을 비교하여 유사성을 검색하고 유사성의 정도를 확률로 제시한다. 또한 객체의 유사성도 객체의 색상, 시공간관계의 특징을 비교하여 유사성을 검색하고 유사성의 정도를 확률로 제시한다. 본 절에서는 유사성의 정도를 표현하기 위한 식을 제시한다.

4.4.1 주석기반검색

주어진 장면과 유사한 장면을 검색하기 위해 사용자에게 의미계층의 장면클래스의 속성인 who, what, where 중 가장 관심 있는 속성을 선택하게 하고 그 속성을 기반으로 유사한 장면의 검색을 수행한다. 이렇게 함으로써 검색공간(search space)을 줄이고 사용자의 기호를 고려하여 사용자가 원하는 검색이 가능하다. 만일 사용자가 관심속성을 선택하지 않는다면 제시된 속성을 or 연산자로 연결하여 장면을 선택한다.

해당하는 속성값을 가진 장면을 검색한 후 그 장면의 객체들과 주어진 장면이 가지는 객체들 사이의 유사성 검색을 시작한다. 유사성은 먼저 의미계층의 객체클래스의 인스턴스인 각 객체의 type과 이름을 먼저 비교한 후 같은 객체가 존재한다면 객체특징계층을 이용하여 특징기반검색을 수행한다.

4.4.2 특징기반검색

주석기반 검색에 의해 같은 객체가 포함되어 있을 경우 객체의 색상, 공간관계를 식별하여 객체가 같은 색상, 같은 공간관계를 가지고 있는지를 검사한다.

(1) 색상의 유사성

색상의 유사성정도는 식 (1)과 같다.

(2) 공간의 유사성

공간의 유사성은 두 객체 사이의 거리에 의해 계산되며, 식은 다음과 같다.

$$Ssi = \sqrt{(x_{q1} - x_1)^2 + (y_{q1} - y_1)^2} + \sqrt{(x_{q2} - x_2)^2 + (y_{q2} - y_2)^2} \quad (2)$$

(단, (x_{q1}, y_{q1}) (x_{q2}, y_{q2}): 검색 대상 장면에서 객체의 좌표, (x_1, y_1) (x_2, y_2): 주어진 장면에서 객체의 좌표, Ssi 는 장면의 i번째 객체의 공간의 유사성 정도)

(3) 정규화(normalize)

색상과 공간의 유사성 정도에 대한 정규화는 다음의 식에 의해 계산된다.

색상의 정규화는 Sim_{col} 을 검정색의 RGB값인 (255,255,255)와 흰색의 RGB값인 (0,0,0)의 차로 나눈다.

$$S_{nci} = \frac{Sim_{col}}{(255-0) + (255-0) + (255-0)} \quad (3)$$

(단 S_{nci} 는 정규화된 색상의 유사성 정도)

공간의 정규화는 S_{si} 를 가장 왼쪽위의 좌표인 (0,0) 과 가장 오른쪽 아래의 좌표인 (9,9)의 거리로 나눈다.

$$S_{nsi} = \frac{S_{si}}{2\sqrt{(9-0)^2 + (9-0)^2}} \quad (4)$$

(단 S_{nsi} 는 정규화된 공간의 유사성 정도)

일반적으로 유사성의 정도는 0일 경우 유사성이 없고 1일 경우 가장 유사한 것으로 간주하므로 1에서 정규화된 값을 뺀값을 유사성 값으로 한다.

$$S_{fnci} = 1 - S_{nci} \quad (5)$$

(단 S_{fnci} 는 색상의 최종 유사성 정도)

$$S_{fnsi} = 1 - S_{nsi} \quad (6)$$

(단 S_{fnsi} 는 공간의 최종 유사성 정도)

4.4.3 장면의 유사성 정도

유사한 장면을 검색할 경우 사용자는 주어진 장면의 배경에 더 많은 관심이 있을 수도 있고 또는, 사람이나 동물과 같은 객체(object)에 관심이 있을 수도 있다. 이런 사용자의 관점을 고려하기 위해 각각의 객체가 유사성에 미치는 영향을 가중치로 표현하고, 장면의 유사성 정도를 계산하기 위해 가중치를 사용한다.

이때 type을 배경과 객체(객체는 배경이외의 모든 유형을 말함)로 구분하고 사용자로부터 각각의 가중치를 입력받고, 배경과 객체에 따른 색상의 가중치와 공간의 가중치를 입력받는다. 예를들어 어떤 사람은 장면 에 나오는 사물이나 인물보다 산이나 강과 같은 배경에 관심이 있을 수도 있고 또 다른 사람은 나무나, 꽃과 같은 사물이나 사람에 관심이 더 있을 수도 있다. 이러한 사용자의 관심을 고려하기 위해 type의 가중치를 사용하고, 한 객체에 대한 관심을 색상과 위치로 구별하여 사용자가 원하는 가장 유사한 객체를 찾을 수 있다. 만일 사용자가 입력하지 않을 경우에는

모든 가중치를 1로 한다.

장면에 있는 i번째 객체의 총 유사성의 정도(total similarity degree)는 다음과 같이 계산된다.

$$S_{ii} = \frac{S_{fnci} * W_{ci} + S_{fnsi} * W_{si}}{W_{ci} + W_{si}} * W_{fi} \quad (7)$$

(단 W_{fi} 는 배경이나 객체에 대한 가중치, W_{ci} 는 색상의 가중치, W_{si} 는 공간의 가중치)

한 장면의 유사성의 정도는 다음의 식에 의해 계산된다.

$$S_{ss} = \frac{\sum_{i=1}^N S_{ii}}{N} \quad (8)$$

(단 N은 주어진 장면에 포함된 객체의 수)

이때 S_{ss} 는 0과 1사이의 값을 갖는다.

주어진 장면과 유사한 장면을 찾은 결과는 (8)의 식에 의해 계산된 S_{ss} 의 값이 큰 순서로 결과를 제시한다.

4.4.4 객체의 유사성 정도

객체의 유사성은 같은 객체번호를 가진 객체에 대해 객체특징계층을 이용하여 특징기반검색을 수행한다. 특징기반 검색은 위에서 제시한 색상, 공간의 유사성과 같으며 객체의 유사성 정도는 다음과 같이 계산된다.

$$S_{oi} = \frac{S_{fnci} * W_{ci} + S_{fnsi} * W_{si}}{W_{ci} + W_{si}} \quad (9)$$

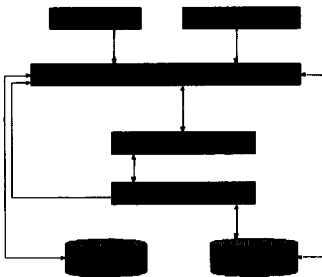
개념계층을 이용한 추석기반 검색과 객체특징계층을 이용한 특징기반 검색을 수행하여 가장 유사한 장면과 객체를 검색할 수 있다. 개념계층을 이용함으로써 검색영역과 검색시간을 줄일 수 있다. 또한 객체특징계층을 통한 검색시 위에서 제안한 유사성 식에 의해 유사성정도를 계산함으로써 사용자에게 결과와 함께 유사성의 정도를 제시할 수 있다.

5. 통합비디오정보시스템 (Hybrid Video Information System : HVIS)

본 장에서는 통합비디오정보시스템을 제안한다. HVIS의 구조를 설명하고, 본 논문에서 제시한 THOMM과 비디오질의언어로 구성할 수 있는 질의 유형과 질의처리 알고리즘을 설명한다.

5.1 HVIS의 구조

통합비디오정보시스템의 구조는 (그림 4)와 같다. HVIS는 사용자 인터페이스, 비디오질의처리기(Video Query Processor : VQP), 상용의 DBMS, 메타데이터베이스, 원시비디오데이터로 구성된다. 사용자는 사용자 인터페이스를 이용해 데이터의 입력과 질의를 수행한다. 메타데이터 입력은 동영상 플레이어에 의해 플레이되는 비디오를 보면서 해당 객체에 속성값을 준다. 특히 색상과 위치는 해당 객체를 사각형의 영역으로 선택함에 의해 절대위치가 계산되어 입력되고, 가장 유사한 색상을 색상테이블에서 선택함에 의해 그 색상의 RGB값이 입력된다. 원시비디오데이터는 프레임의 집합으로 구성된 디지털비디오이고, DBMS는 상용의 데이터베이스관리자로서 메타데이터베이스를 검색한다.



(그림 4) HVIS의 구조

비디오질의처리기는 사용자로부터 질의를 입력받아 질의를 분석하는 질의분석기와 상용의 DBMS에서 사용되는 SQL문을 생성하는 질의어 생성기로 구성된 전처리기이다. 상용의 DBMS는 해당하는 질의의 결과를 찾기 위해 메타데이터베이스를 검색하고 검색결과를 플레이한다.

VQP의 처리단계는 다음과 같다.

- 단계 1 : 스캐너에 의해 Lexical Analysis과정 수행 수행결과 토큰을 생성한다.
- 단계 2 : Parser에 의해 Syntax Analysis 수행 수행결과 parse tree를 생성한다.
- 단계 3 : parse tree를 입력으로 SQL문장 생성
- 단계 4 : 메타데이터베이스 탐색

위의 처리과정에 의해 생성된 SQL문은 상용의 DBMS에 의해 검색을 수행한다. 만일 검색결과가 존재한다면

면 검색결과가 제시되고 원하는 결과를 선택하면 동영상 플레이가 실행되어 결과 영상을 플레이한다.

5.2 질의유형

제안된 VQL에 의해 다음과 같은 질의를 구성할 수 있다.

- ① TYPE 1 : 비디오 질의
: 비디오질의는 사용자가 지정하는 비디오를 검색한다.
- ② TYPE 2 : 구조질의
: sequence, scene에 해당하는 비디오 구조를 통해 검색한다.
- ③ TYPE 3 : 객체질의
: 장면에 등장하는 인물이나, 배경등과 같은 관심 있는 객체를 검색하는 것이다.
- ④ TYPE 4 : 객체의 색상질의
: 객체의 색상질의는 개체의 색상과 관련된 질의로 RGB값을 질의 조건으로 입력받거나, 색상을 직접 단어로 입력받은 후 객체특징(object_feature)클래스를 검색하는 것이다.
- ⑤ TYPE 5 : 객체의 시간관계질의
: 하나의 장면내에서 객체들간의 시간관계를 질의조건으로 하여 객체특징(object_feature) 클래스를 검색한다.
- ⑥ TYPE 6 : 객체의 공간관계 질의
: 객체에 대한 절대 공간관계 질의와 상대 공간관계 질의가 가능하며 질의를 입력받아 객체특징(object_feature) 클래스를 검색한다.
- ⑦ TYPE 7 : 유사 질의
: 주어진 장면과 유사한 장면이나, 주어진 객체와 유사한 객체를 검색하는 유사질의는 THOMM의 개념계층을 이용한 주석기반 검색과 객체특징계층을 이용한 특징기반 검색을 수행하여 가장 유사한 장면과 객체를 검색한다. 4.4절에서 제안한 유사성식에 의해 계산된 유사성 정도와 함께 결과를 제시한다.

5.3 질의처리 알고리즘

시스템은 먼저 시스템을 사용하는 사용자에게 사용자가 관심 있는 장르와 분야, 주제를 입력받아 지식베이스(knowledge_base)에 사실(fact)로 저장된다. 질의 결과가 많을 경우 이 지식은 사용자의 질의처리시 filtering을 위해 사용된다. 즉, 질의어 생성시 사용자의

관심장르와 관심분야를 질의어에 포함시킨다. 사용자의 질의는 주석기반질의, 특징기반질의, 유사질의로 구성될 수 있으며, 이 질의어를 입력받아 질의분석을 수행하고, 질의어를 생성한 후 상용의 DBMS에 의해 검색을 수행한다. 만일 검색결과가 존재한다면 검색결과를 플레이한다.

TYPE 1, 2, 3의 질의는 SQL문장과 일대일 대응하며, TYPE 4, 5, 6은 다음과 같은 문장으로 변환되어 상용의 DBMS에 의해 검색된다.

```

if (type 4 질의)
  if (색상)
    mapping_color(RGB) 호출
    return(RGB);
  select s1.frame_duration
  from object_feature s, object o
  where (object조건질의) and (s.o_id=o.o_id) and color_fun
  (s.color.RGB) <= ε (임계치=30)
if (type 5질의)
  select s1.frame_duration, s2.frame_duration
  from object_feature s1, object_feature s2, object o1, object o2
  where (object o1 검색조건) and (object o2 검색조건) and
  (s1.o_id=o1.o_id) and (s2.o_id=o2.o_id) and (s1.scene_id=
  s2.scene_id) and temporal (s1.frame_duration, s2.frame_
  duration, 시간검색조건)
if (type 6 질의)
  if (절대공간질의)
    select s.frame_duration
    from object_feature s, object o
    where (객체검색조건) and (o.o_id=s.o_id) and
    position(s1.position, s2.position, 공간관계검색조건)
  if (상대공간질의)
    select s1.frame_duration, s2.frame_duration
    from object_feature s1, object_feature s2, object o1, object o2
    where (object검색조건) and (object검색조건) and (s1.o_id=
    o1.o_id) and (s2.o_id=o2.o_id) and (s1.scene_id=
    s2.scene_id) and spatial(s1.position, s2.position, 공간검
    색조건)
  
```

유사성 질의를 위한 질의처리 알고리즘은 다음과 같다. 유사성 질의처리 결과는 유사성정도가 큰 순서로 제시된다.

```

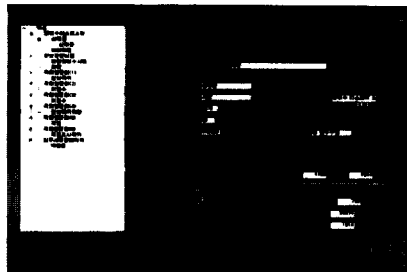
if (type 7 질의)
  유사성 결과 리스트=∅
  Sim=0
  if (유사한 장면질의)
  {
  for all (주석기반에 의해 선택된 장면)
  for all (주어진 장면의 객체)
  { if (선택된 장면에 type과 name이 같은 객체가 존재)
  //색상의 유사성검색
  Sci=R0-R1+G0-G1+B0-B1
  Snci =  $\frac{S_{ci}}{(255-0) + (255-0) + (255-0)}$ 
  
```

```

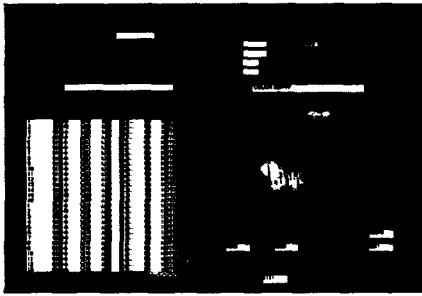
Sfnrci = 1- Snci
//공간의 유사성검색
Ssi =  $\sqrt{(x_{a1}-x_1)^2 + (y_{a1}-y_1)^2} + \sqrt{(x_{a2}-x_2)^2 + (y_{a2}-y_2)^2}$ 
Snsi =  $\frac{S_{si}}{2\sqrt{(9-0)^2 + (9-0)^2}}$ 
Sfnsci = 1- Snsi
//한 객체의 총 유사성
if (객체의 type이 배경)
  Sti =  $\frac{S_{fnrci} * W_{rci} + S_{fnsci} * W_{sci}}{W_{rci} + W_{sci}} * W_{ti}$ 
else
  Sti =  $\frac{S_{fnrci} * W_{rci} + S_{fnsci} * W_{sci}}{W_{rci} + W_{sci}} * W_{ti}$ 
Sim=Sim+Sti
}
Sim_deg=Sim/N
}
else (객체의 유사성 질의)
// 장면의 유사성 중 한 객체의 공간의 유사성과 색상의 유사
성 알고리즘은 위와 같다.
Sim_deg =  $\frac{S_{fnrci} * W_{rci} + S_{fnsci} * W_{sci}}{W_{rci} + W_{sci}}$  //객체의 최종 유사성
if (Sim_deg >= τ)
  유사성 결과 리스트= 유사성 결과 리스트 U Scene_id
}
  
```

6. 구 현

본 논문에서 제안한 THOMM과 비디오질의언어에 기반을 둔 통합비디오정보시스템을 TV 뉴스를 도메인으로 구현했다. 사용자는 해당하는 비디오를 보면서 입력할 수 있고, 비디오와, 시퀀스, 장면이 선택되면 해당 속성들이 제시되고 제시된 속성에 속성값을 입력하면 된다. 객체에 대한 입력중 객체의 색상은 색상테이블에서 가장 유사한 색을 선택하여 입력할 수도 있고 직접 색상을 입력할 수도 있다. 객체의 위치는 비디오의 객체를 마우스를 이용하여 관심있는 객체의 위치를 사각형으로 영역을 설정하여 클릭함에 의해 자동적으로 절대 주소가 계산되어 입력된다. 객체에 대한 입력화면은 (그림 5)와 같다. 객체에 대한 질의결과 화면은 (그림 6)과 같다.



(그림 5) 입력 화면



(그림 6) 질의결과 화면

7. 결론 및 향후 연구과제

본 논문에서는 특징기반 검색과 주석기반 검색을 통합하여 다양한 사용자의 의미검색을 지원하고 유사성 질의를 지원하는 통합비디오정보시스템(HVIS)을 제안했다.

HVIS에서는 한편의 비디오를 비디오다큐먼트, 시퀀스, 장면, 객체로 나누고 물리적인 비디오스트림을 위한 원시비디오데이터계층(rawvideo-data layer)과 주석기반 검색, 특징기반 검색, 유사성 검색을 지원하기 위한 메타데이터계층(meta-data layer)의 두개의 계층을 가진 통합 객체지향 메타데이터모델(THOMM)과 이 모델을 기반으로 주석기반 질의, 특징기반 질의, 유사질의가 가능한 비디오질의언어와 질의를 처리하기 위한 비디오질의처리기와 질의처리 알고리즘, 유사성의 정도를 확률로 표시하기 위한 유사성식을 제안했다. HVIS는 주석기반검색과 특징기반검색을 통한 사용자의 다양한 의미검색을 지원하며, 특히 유사성 질의처리시 개념계층을 이용한 주석기반 질의를 먼저 수행함으로써 검색영역과 검색시간을 줄일 수 있고, 객체특징계층을 이용한 색상과 공간의 유사성 계산식에 의해 유사성의 정도를 제시한다. 비디오 질의처리기(VQP)는 구조질의, 객체질의,객체의 시공간질의와 유사성질의의

문장을 받아들여 상용의 DBMS에서 사용할 수 있는 SQL문을 생성하며, 특히 유사성 질의처리시 유사성의 정도를 확률로 표현할 수 있다. 유사성식은 사용자의 관심에 따라 색상, 공간, 객체의 유형에 가중치를 부여하여 사용자가 원하는 관점의 가장 유사한 장면, 객체의 검색이 가능하다. 즉, 어떤 사용자는 장면에 나오는 사람이나, 꽃과 같은 객체보다는 배경에 더 많은 관심을 가질 수 있다. 유사성 검색시 사용자가 배경의 가중치에 다른 것보다 큰 값을 부여하면 검색시 배경이 같은 것이 더 큰 유사성 확률을 갖게되고, 객체의 공간과 색상의 가중치를 부여함으로써 사용자가 관심을 가진 속성의 값이 가장 유사한 객체가 높은 확률 값을 갖게된다. 검색된 결과는 확률이 높은 순서로 제시되고 원하는 결과를 사용자가 클릭하면 동영상플레이어에 의해 결과가 플레이된다. 이렇게 유사성식을 사용하여 유사한 많은 결과를 확률과 함께 제시함으로써 사용자는 제시된 확률값을 보고 원하는 가장 유사한 장면을 쉽게 선택할 수 있다.

HVIS가 기존의 비디오데이터베이스 시스템과 다른 점은 주석기반검색과 특징기반검색을 모두 지원한다. 즉 주석기반의 객체질의 사건질의가 가능하며, 특징기반의 색상질의 시간질의 공간질의가 가능하다. 둘째로 장면이나 객체에 대한 유사성질의가 가능하다. 유사성질의는 주석기반질의에 의해 검색된 장면이나 객체에 대해 특징기반 검색을 수행하게된다. 그러므로 탐색의 영역을 줄일 수 있다. 또한 유사성의 정도를 확률로 표시하여 사용자에게 제시할 수 있다. 특히 유사한 객체나 장면을 찾기 위해 가중치를 사용하여 유사성의 정도를 확률로 표현함으로써 사용자가 가장 원하는 객체와 장면에 대한 검색이 가능하다. 본 논문에서 제시한 HVIS와 기존의 다른 시스템과의 비교는 <표 2>와 같다.

향후 연구과제로는 효율적인 검색을 위한 인덱싱기법과 불완전한 사용자의 질의를 처리하기 위한 지식구

<표 2> HVIS와 기존의 시스템과의 비교

질의유형	시스템	VENUS	OVID	AVIS	KMeD	HVIS
객체기반질의		가능	가능	가능	가능	가능
사건기반질의		불가능	가능	가능	불가능	가능
색상질의		불가능	불가능	불가능	불가능	가능
유사성질의		불가능	불가능	불가능	가능	가능
질의언어		CVQL	not exist	not exist	P-Query	VQL
분 석		특징기반질의	주석기반질의	주석기반질의	주석기반질의, 공간질의 가능, but 의료분야의 이미지 데이터에 대해서만 적용가능	주석기반질의, 특징기반질의, 시공간질의, 색상질의, 유사성질의, 특히 유사성의 정도를 확률로 표시

성 방법, 질의 완화기법과 어떻게 그 질의결과가 생성되었는지를 제시할 수 있는 협동적인 질의 응답(cooperative query answering) 기법의 연구가 필요하다.

참 고 문 헌

[1] T.C.K Kuo, and Arbee L.P.Chen, "Indexing, Query interface and Query Processing for Venus: A Video Database System," <http://piggy.cs.cthu.edu.tw>, Taiwan, 1996.

[2] R. Hjesvold, "VIDEO-Star-A Database for Video Information Sharing," Ph.D. Thesis, Norwegian Institute of Technology, Nov.1995.

[3] Kingsley C. Nwosu, Bhavani Thuraisinghan, P. Bruce Berra, "Multimedia Database Systems - A New Frontier," IEEE Multimedia database Systems, pp.21-23, 1997.

[4] P. Senthil Kumar, G. Phanendra Babu, "Intelligent multimedia data : data indices+inference," Multimedia Systems, Vol.6, pp.395-307, 1998.

[5] S.W. Smoliar and H. Shang, "Content-Based Video Indexing and Retrieval," IEEE Multimedia, Summer, 1994.

[6] Michail Vazirgiannis, Yannis Thodorids, Timos Sellis, "Spatio-temporal composition and indexing for Large Multimedia Applications," Multimedia Systnes, Vol.6, pp.284-298, 1998.

[7] Wesley W. Chu, Alfonso F. Cardenas, and Ricky K. Taira, "KMED : A Knowledge-Based Multimedia Medical Distributed Database System," Information Systems, Vol.20, No.2, pp.75-96, 1995.

[8] S. Adali, K.S.Candan, S.S.Chen, K.Erol, and V.S.Subrahmanian, "Advanced Video Information System," ACM Multimedia Journal, 1995.

[9] Eitetsu Omoto and Katsumi Tanaka, "OVID : Design and Implementation of a Video-object Database System," IEEE trans. on knowledge and data engineering, Vol.5, No.4, August 1993.

[10] J. Monaco, "How to Read a Film. The Art, Technology, Language, History and Theory of Film and Media," Oxford University Press, 1981.

[11] John Z. Li, M. Tamer Oisu, and Duane Szafron, "Querying Languages in Multimedia Database

Systems," TR95-12, The university of Alberta, Canada, 1996.

[12] W.I.Grosky, "Multimedia Information Systems," IEEE Multimedia, Vol.1, No.3, Fall 1994.

[13] J. F. Allen, "Maintaining knowledge about temporal intervals," Communications of ACM, Vol.26, No.11, pp.832-943, 1983.



윤 미 희

e-mail : ymh@cs.sookmyung.ac.kr
 1990년 숙명여대 전산학과 졸업 (학사)
 1992년 숙명여대 대학원 전산학과 (석사)
 1999년 숙명여대 대학원 전산학과 (이학박사)

관심분야 : 멀티미디어, 지식베이스, 멀티미디어 데이터 베이스 등



윤 용 익

e-mail : yiyoon@cs.sookmyung.ac.kr
 1983년 동국 대학교, 통계학과
 1985년 한국과학기술원 전산학과 (석사)
 1994년 한국과학기술원 전산학과 (박사)

1985년~1997년 한국전자통신연구원 책임연구원
 1997년~현재 숙명여대 전산학과 교수

관심분야 : 분산 시스템, 실시간 시스템, 컴퓨터 통신과 네트워크 등



김 교 정

e-mail : kiokchkim@sookmyung.ac.kr
 1972년 연세대학교 화학과 졸업(학사)
 1983년 Clarkson Univ. 전산학과 (석사)
 1991년 Clarkson Univ. 전산학과 (박사)

1986년~현재 숙명여대 전산학과 교수
 관심분야 : 지식베이스, 데이터마이닝 등