

의미 기반 주석을 이용한 비디오 검색 시스템의 설계 및 구현

홍 수 열*

Design And Implementation of Video Retrieval System for Using Semantic-based Annotation

Soo Youl Hong *

요 약

비디오는 broadcasting, 교육, 출판과 군사 등 다양한 응용들과 함께 멀티미디어 컴퓨팅과 통신 환경의 중요한 요소가 되었다. 멀티미디어 데이터 검색을 위한 효과적인 방법의 필요성은 대용량의 멀티미디어 응용들에서 날로 증가하고 있다. 따라서, 비디오 데이터의 검색과 표현은 비디오 데이터베이스에서 주요 연구 이슈 중에 하나가 되었다. 비디오 데이터의 표현 방법으로 주로 2가지 접근 방법이 있다: (1) 내용 기반 비디오 검색 과 (2) 주석 기반 비디오 검색.

이 논문은 의미 기반 주석을 이용한 비디오 검색 시스템을 설계하고 구현한다.

Abstract

Video has become an important element of multimedia computing and communication environments, with applications as varied as broadcasting, education, publishing, and military intelligence. The necessity of the efficient methods for multimedia data retrieval is increasing more and more on account of various large scale multimedia applications. Accordingly, the retrieval and representation of video data becomes one of the main research issues in video database. As for the representation of the video data there have been mainly two approaches: (1) content-based video retrieval, and (2) annotation-based video retrieval.

This paper designs and implements a video retrieval system for using semantic-based annotation.

* 동주대학 컴퓨터 정보통신 계열 조교수
본 연구는 동주대학 학술연구조성비에 의해 연구되었음.

I. Introduction 소개

불과 몇 년 전만 해도 시각적 매체의 검색과 표현 문제는 특별한 이미지 데이터베이스(지리학, 의학, 비행모의 훈련 등)와 시청각 산업(방송과 공문서 보관소) 등의 전문적인 응용과 컴퓨터화된 훈련, 교육 등에 국한되어 왔다.

최근에 컴퓨터 하드웨어의 발달로 일상 생활에서 사용하는 데이터가 텍스트 중심의 데이터에서 그래픽 중심의 시각적인 데이터가 많이 사용되고 있다.

그래픽 중심의 데이터 중에서 비디오 데이터가 차지하는 비중이 높고 아주 중요한 데이터 요소가 된다. 이런 비디오 데이터는 대용량의 데이터로서 다루기 힘든 데이터 자원이기 때문에 효과적으로 데이터를 관리할 수 있는 비디오 데이터베이스와 비디오 데이터를 효과적으로 접근 할 수 있는 인덱싱에 대한 연구가 활발하게 전개되고 있다.

비디오 데이터베이스와 인덱싱을 위한 비디오 데이터의 표현에 관한 중요한 두 가지 접근 방법이 있다.

첫 번째 접근 방법은 내용 기반 검색(Content-based retrieval)이다. 내용 기반 검색은 비디오의 특징(Feature)을 이용하여 비디오내의 특징을 추출하여, 그 특징들을 인덱싱을 하는 접근 방법이다.

비디오 특징을 나타내는 종류에는 객체(Object)의 움직임, 카메라의 움직임, 색의 분포 등 다양하게 있고, 보다 정확한 검색을 위해서는 이미지에 대한 전반적인 이해가 요구되며, 이미지 프로세싱 기법에 대한 지식이 요구된다. 아직까지는 초보단계의 연구가 진행되고 있으며, 많은 연구가 필요하다(1).

두 번째 접근 방법은 주석 기반 검색(Annotation-based retrieval)이다. 주석 기반 검색은 사용자가 비디오 데이터를 일일이 보면서 특징에 맞는 내용을 주석으로 붙이거나, 자동화 된 도구(Tool)를 이용하여 사용자가 주석을 붙이는 방법 등으로 나눌 수 있다. 이 검색 방법은 수동적인 작업으로 인하여 많은 작업량이 발생되지만 구현하기가 쉽다(2).

따라서 본 논문에서는 데이터베이스 내에 저장되어 있는 비디오 데이터를 검색하는 방법으로 비교적 접근하기 쉬운 주석 기반 검색 중 비디오 데이터의 의미(Semantics)를 주석으로 붙이는 의미 기반 주석(Semantic-based Annotation) 비디오 검색 방법을 설계하고 구현 할 것이다.

II. 비디오 검색 방법

2-1. 내용 기반 검색(Content-based retrieval)

일반적으로 내용기반 검색은 비디오의 특징(Features)을 추출하고, 추출된 특징에 대한 내용을 비디오 데이터에 인덱싱(Indexing)하여 사용자가 텍스트(Text)나 비주얼 이미지 데이터에 기반 한 검색을 통해 DB를 접근하거나 의미 있는 아이콘의 표현과 상호 작용을 통해 브라우저(Browse)할 수 있는 비디오 검색 방법이다(1).

내용기반 검색의 비디오 데이터 관리 구조의 블록도는 그림 1과 같다.

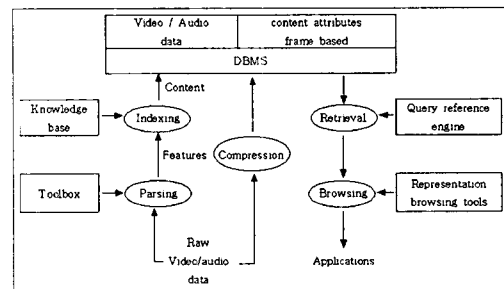


그림 1. 비디오 관리 구조 블록도
Fig 1. Diagram of video management architecture

이 내용기반 검색을 위한 시스템은 크게 4부분으로 구성된다.

◎ 비디오 분석(Parsing)

비디오 분석은 원본 데이터(Source material)의 비디오 스트림(Video stream)을 의미적 프리미티브(Semantic Primitives)인 클립(Clips)으로 분할하는 단계를 말하며, 이 들 프리미티브를 기초로 한 비디오

클립을 기존의 지식 표현 기술을 사용하여 의미적 표현 (Semantic Description)에 의한 인덱스를 만들 수 있다.

◎ 인덱싱(Indexing)

비디오 분석(Parsing)에서 얻어진 비디오 클립들을 태그(Tag)와 함께 데이터베이스에 등록하는 단계이다. 태그는 비디오 이미지들의 의미적 프리미티브에 의해 분류를 제어하는 지식 모델에 기반 한 정보를 포함한다. 인덱싱은 이미지 자체에 의해 운용되고, 의미적 표현은 모델에 의해 제공된다.

◎ 추출과 브라우징(Retrieval and Browsing)

사용자가 텍스트(Text)나 시각적 예제(Visual examples)에 기반 한 조회(Queries)를 통해 DB를 접근하거나 또는 의미 있는 아이콘의 표현과 상호 작용을 통해 브라우저 한다. 또한 사용자는 추출 조회(Retrieval query)의 결과를 브라우저 할 수도 있으며, 추출과 브라우징 모두 사용자의 시각에 호소한다는 것이 중요하다.

2.2 주석 기반 검색

(Annotation-based retrieval)

주석 기반 검색의 편집 시스템은 비디오 데이터를 하나 혹은 그 이상의 비디오 세그먼트(Segments)들로 나누고, 각 세그먼트로 나누어진 비디오 데이터를 재생시키면서 주석을 작성하기 위한 특정 구간으로 시작 프레임 번호와 끝 프레임 번호를 결정한다. 비디오 데이터의 특정 구간에 대해 (속성, 값)으로 구성된 주석 원소들을 지정하고, 속성은 주석의 의미를 설명하며, 값은 실제 내용을 표현하는 것으로 '속성=값'이란 의미를 갖게 한다. 지정된 주석 원소들은 제목, 프레임 구간 정보 등의 비디오 데이터 정보와 함께 하나의 비디오 단위로 DB에 저장 관리한다.

주석 기반 검색은 위와 같이 작성된 주석 DB를 이용하여 사용자가 원하는 비디오 데이터를 검색하는 과정이다. 먼저 특정 비디오 또는 전체 비디오 등과 같이 검색 범위를 결정하고 사용자가 원하는 비디오 검색 조건을 사용하여 질의(Query)를 하며, 검색 범위 내에서 모든 검색 조건들을 만족하는 비디오 단위들을 결과로 나타낸다(8).

주석 기반 검색 시스템의 구조는 그림 2와 같다.

◎ 의미적 관계 질의(Semantic Relationship Query)

비디오 검색 시스템은 사용자들이 기본적인 질의와 관계 질의, 접속어 질의 등을 포함하는 질의를 만들도록 되어 있다.

특히 의미적 관계 질의는 공통 속성 값이 의미적 관계가 있는 비디오 단위에 존재하는지를 질의하는 것이다.

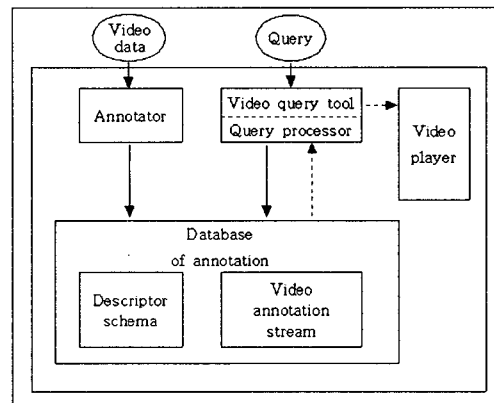


그림 2. 주석 기반 검색 시스템의 구조
Fig 2. Structure of annotation-based video retrieval system

Ⅲ. 의미 기반 주석(Semantic-based Annotation) 비디오 검색 시스템

3.1 시스템 설계

본 논문에서 제안된 의미 기반 주석은 주어진 비디오 (또는 이미지)의 의미(Semantic)란 실제 누가 무엇을 행(Action)하고 있거나, 어떤 상태에 있음을 기술할 수 있다. 즉 비디오 데이터에는 행위자(Actor)가 있고, 그 행위자가 무엇을 행하고 있거나 어떤 상태에 있음(Action 또는 State)을 기술함으로써 비디오 데이터의 의미(Semantic)를 어느 정도 명확하게 할 수 있다. 따라서 본 논문에서는 비디오 데이터의 의미(Semantic)를 주석으로 붙이는 비디오 검색 방법을 제안한다.

그림 3은 의미 기반 주석 비디오 검색 시스템의 구조이며 시스템의 설계를 위한 각 단계를 아래에서 기술한다.

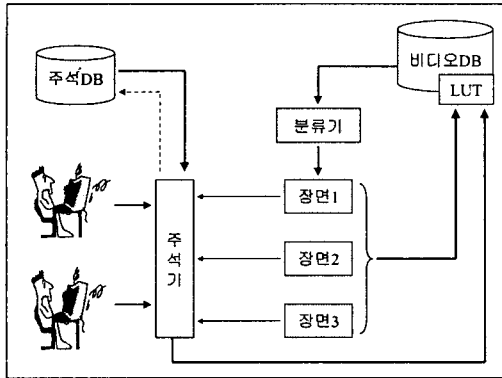


그림 3. 의미 기반 주석 비디오 검색 시스템의 구조
Fig 3. Structure of semantic-based annotation video retrieval system

첫 번째 단계로 연속적인 프레임들의 흐름인 비디오 스트림(Video stream)을 하나, 하나의 장면(Shot)들로 분류하는 단계이다. 이것은 사용자가 비디오를 보면서 의미 있는 장면들로 분류 가능하지만 작업량을 줄이기 위해 자동화된 분류를 수행할 수도 있다.

두 번째 단계로 분류된 장면들에 대해 사용자가 행위자와 행위를 직접 기술하는 주석 기술 단계이다. 주목하는 장면을 보면서 사용자가 주석을 붙이게 된다. 이때 공통된 주석을 붙이기 위해 범위를 정해야 하고 주석은 그림 4와 같이 정의된 범위 내에서 행위자와 행위를 상위 단계로 하는 계층적인 구조를 보이게 한다.

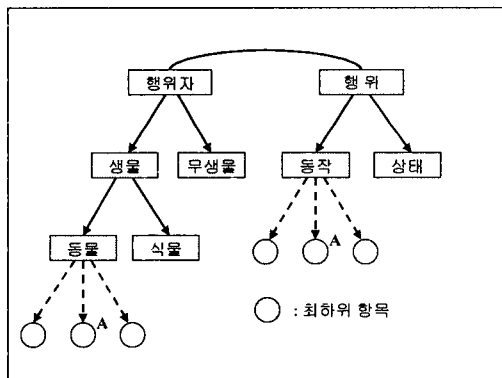


그림 4. 주석의 구조
Fig 4. Structure of annotation

실제 비디오 데이터에서 행위자는 크게 사람과 같은 동적 움직임이 있는 것(생물)과 빌딩과 같은 동적 움직임이 없는 것(무생물)으로 나눌 수 있다. 예를 들면 아

나운서가 뉴스를 진행하고 있는 경우에 사용자는 '행위자 → 생물 → 동물'이라는 계층의 하부계층인 사람이 말을 하고 있는 상태(또는 동작으로 볼 수 있음) 장면으로 '동작 → 말하다'로 주석을 붙일 수 있다.

```

Video_stream { //비디오 스트림
    Video_stream_file name;
    Video_stream_annotation;
    Video_stream_frame#;
    Shot:
}
Shot { //장면
    Previous_shot#:
    Key_frame_number;
    End_frame_number;
    Shot_annotation;
    Next_shot#:
    Annotation:
}
Annotation { //주석
    annotation_keyframe_number;
    annotation_frame_number;
    actor;
    action;
    when;
    how;
    where;
}
    
```

그림 5. 주석의 표현
Fig 5. Description of annotation

또는 빌딩이 있는 장면에서는 '행위자 → 무생물'이라는 계층에서 하부계층으로 빌딩이 있는(상태) 장면으로 주석을 간단하게 붙일 수 있다. 이러한 주석은 정의되어 있는 범위 내에서 붙여져야 하고 만일 정의된 것이 없을 경우 사용자가 필요한 부분을 추가할 수 있다. 이때 비디오 데이터의 의미에서 행위자에 대한 객체(Object)가 존재할 수 있는 경우도 있으며 그림 5의 상위 계층에 객체를 추가하여 확장이 가능하다. 실질적으로 이것은 정확하게 '누가, 언제, 어디서, 어떻게, 무엇을 하겠다'라고 하는 6가지 범위까지 확장 가능하다. 또한, 이미 주석이 정의된 유사한 장면이 있는 경우 그 주석을 그대로 이용할 수 있다.

그림 5는 주어진 장면에 대한 주석의 표현으로 Video_stream과 Shot의 끝 부분인 '#'을 일련의 비디오 스트림과 비디오 스트림 내의 Shot의 일련번호를 의미한다.

비디오 검색을 위해 사용자는 주석의 내용을 질의로

써 입력할 수 있게 한다. 질의를 위한 주석의 종류에는 어떤 종류의 비디오 데이터를 선택할 것인가를 질의하는 비디오 스트림 주석과 비디오 데이터의 장면에 주석을 붙여 장면의 의미를 찾는 장면 주석(Shot annotation), 그리고 각 장면마다 행위자, 행위, 언제, 어떻게, 어디서와 같은 의미를 붙여 각 장면의 상태를 검색할 수 있게 한다.

이와 같이 비디오 데이터에 의미를 주석으로 표현하며, 비디오 데이터를 검색할 수 있게 하는 것이 본 논문의 요지가 된다.

3.2 비디오 검색 편집 시스템의 구현

본 논문 3.1절에서 제안한 내용을 중심으로 비주얼 베이직 언어와 데이터베이스(Access)를 사용하여 개인용 컴퓨터에서 구현하였다. 의미 기반 주석을 이용한 비디오 편집 시스템은 비디오 검색을 위해 비디오 데이터와 주석 데이터를 DB에 등록하는 프로그램으로 아래와 같이 세 부분으로 구분된다. 첫째 비디오 데이터 검색에 필요한 비디오 데이터 파일과 비디오 데이터의 내용을 주석으로 등록하는 비디오 스트림 등록 부분이다.

둘째, 등록된 비디오 스트림을 장면별로 나누어 등록하는 장면 구분 등록 부분이다.

셋째, 장면 별로 나누어진 비디오 데이터에서 부분별로 의미 있는 주석을 등록하는 부분이다.

실제 구현된 프로그램의 화면은 그림 6과 같다.

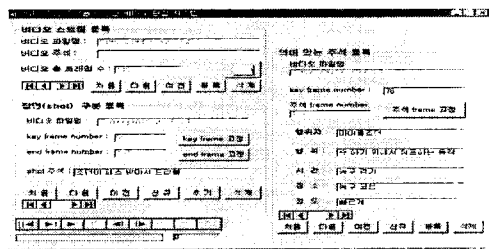


그림 6. 의미 기반 주석을 이용한 비디오 편집 시스템
Fig 6. Video edition system for using semantic-based annotation

3.2.1 비디오 스트림 등록

비디오 스트림 등록 부분은 여러 종류의 비디오 데이터 파일을 등록하는 부분이다. 이 비디오 데이터 파일은 윈도우즈에서 사용 가능한 모든 압축 파일을 사용할 수 있다. 비디오 스트림 등록에서 나타난 각 항목들의 기능을 살펴보자.

항 목	기 능
· 비디오 파일명	비디오 데이터 파일명 등록
· 비디오 주석	비디오 데이터 내용을 주석으로 입력
· 비디오 프레임수	비디오 데이터의 총 프레임 수가 나타남
· 처음	등록된 비디오 데이터 중 맨 처음 비디오 데이터를 검색
· 다음	현재 화면상에 나타난 다음 비디오 데이터를 검색
· 이전	현재 화면상에 나타난 이전 비디오 데이터를 검색
· 등록	새로운 비디오 데이터를 입력하여 등록
· 삭제	현재 화면상에 나타난 비디오 데이터를 삭제

3.2.2 장면(Shot) 구분 등록

비디오 스트림 등록에 등록된 비디오 데이터를 장면(shot) 별로 나누어, 각 장면에 맞는 의미 있는 주석을 입력하여 등록한다.

장면 구분 등록에서 나타난 각 항목들을 살펴보자.

항 목	기 능
· 비디오 파일명	비디오 스트림에서 등록되어 나타난 파일명
· 키 프레임 번호	장면에서 핵심이 되는 화면의 번호가 나타남
· 키 프레임 고정	키 프레임 번호를 지정하는 기능을 함
· 마지막 프레임 번호	장면의 마지막 프레임이 되는 번호가 나타남
· 마지막 프레임 고정	마지막 프레임 번호를 지정하는 기능을 함
· 장면 주석	비디오 데이터를 장면별로 분할했을 때, 장면의 핵심이 되는 내용을 주석으로 등록
· 처음	현재 나타난 비디오 파일의 장면 중 맨 처음에 등록된 장면의 내용을 검색하여 나타남
· 다음	현재 화면상에 나타난 장면의 다음 장면의 내용을 검색하여 나타남
· 이전	현재 화면상에 나타난 장면의 이전 장면의 내용을 검색하여 나타남
· 신규	현재 나타난 비디오 파일의 맨 처음 장면을 등록할 때 사용함
· 추가	새로운 장면을 추가 할 때 사용
· 삭제	현재 화면상에 나타난 장면을 삭제할 때 사용

3.2.3 주석(Annotation) 등록

주요 장면 별로 나누어진 비디오 데이터의 각 프레임 별로 행위자, 행위, 시간, 장소, 정도 등의 내용을 주석으로 상세하게 기록하는 부분이다.

주석 등록에서 나타난 각 항목들의 기능을 살펴보자.

항목	기능
· 주석 frame 고정	주석을 붙일 장면을 선택하는 기능
· 행위자	장면의 주체 등록
· 행위	장면의 주체가 하는 행위 등록
· 시간	장면에서 나타난 시간 등록
· 장소	장면에서 나타난 장소 등록
· 정도	장면에서 어떻게, 얼마만큼 행하고 있는지의 정도 등록

3.3 비디오 검색 시스템의 구현

본 논문에서 제안된 비디오 검색 시스템은 비디오 편집 시스템에서 등록된 비디오 데이터와 검색을 위한 의미 기반 주석 데이터를 사용자가 쉽게 검색할 수 있게 설계된 시스템이다. 비디오 검색을 위해 사용자는 주석의 내용을 행위자, 행위, 시간, 장소, 정도 등 5가지 형태의 질의어로 입력할 수 있다. 입력된 내용과 가장 많이 질의하는 항목들을 주석으로 화면상에 보여주고, 사용자가 보여진 주석 중에서 마우스를 선택하여 재생 단추를 누르면 해당 비디오 데이터가 디스플레이 된다.

본 논문에서 구현된 비디오 검색 시스템 구현 화면은 그림 7과 같다.

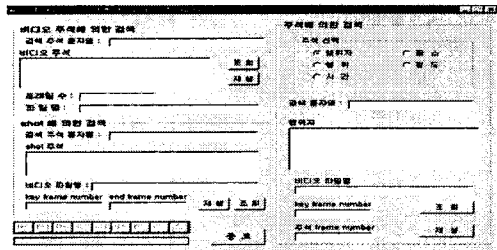


그림 7. 의미 기반 주석을 이용한 비디오 검색 시스템
Fig 7. Video retrieval system for using semantic-based annotation

◎ 비디오 주석에 의한 검색

여러 종류의 비디오 데이터가 존재할 때, 사용자가 원하는 비디오의 제목이나 의미를 입력하여 조회 단추를 누르면 입력된 질의어와 유사한 종류의 항목들이 화면상에 나타난다. 그 항목들 중에서 사용자가 원하는 항목을 선택하며 재생 단추를 누르면 해당 비디오 데이터가 화면상에 디스플레이 된다.

◎ 장면(shot)에 의한 검색

여러 종류의 비디오 데이터가 존재할 때, 사용자가

원하는 비디오의 장면들을 여러 종류의 비디오에서 수집하여 비교 분석하고자 할 때 편리한 검색방법이다. 예를 들면 바다가 있는 장면들을 찾고자 할 때 검색 문자열에서 '바다'를 입력하면 여러 종류의 비디오데이터에서 바다와 관련된 항목들이 화면상에 나타나게 될 것이다. 이들 중 원하는 항목을 선택하여 재생 단추를 누르면 해당 비디오 데이터가 화면상에 디스플레이 된다.

◎ 주석에 의한 검색

주석에 의한 검색 방법은 사용자가 비디오에서 어떤 영화배우가 어떤 영화에 출연했는지, 어떤 동작이 어떤 영화에서 나타났는지, 어느 시대의 영화인지, 어떤 장소가 어떤 영화에서 배경으로 나타났는지, 어느 정도의 긴급한 상황이 어떤 영화에서 나타났는지 등의 검색이 필요할 때 주석 선택에서 5가지 중 1가지를 선택하고 검색문자열에 찾고자 하는 의미 문자열을 입력하고, 조회 단추를 누르면 관련된 항목들이 화면상에 나타난다.

이들 중 원하는 항목을 선택하여 재생 단추를 누르면 해당 비디오 데이터가 화면상에 디스플레이 된다.

IV. 결론 및 향후 과제

본 논문은 비디오 데이터의 의미를 주석으로 이용한 비디오 검색 편집 시스템과 비디오 검색 시스템을 구현하였다. 이러한 방법은 수동적인 작업으로 인하여 많은 작업량을 요구하지만 구현이 용이하다는 장점을 가지고 있다. 따라서 작업량을 최소화할 수 있는 자동화 도구(Tools)의 연구가 요구된다.

본 논문은 모든 것이 PC에서 소프트웨어로 구현되었지만 VOD(Video-On-Demand), 비디오 CD, 영화 예편 등의 제작이나, 중요한 장면 분류 및 검색과 같은 많은 곳에서 응용될 수 있을 것으로 기대된다. 향후 연구 과제로는 주석 기반 검색의 주석을 수작업으로 하는 대신 자동으로 할 수 있는 자동화 도구 개발에 많이 투자할 것이다.

참고문헌

- [1] Stephen W. Smoliar, HongJiang Zhang, "Content-Based Video Indexing and Retrieval", IEEE, 1994.
- [2] Ki-Wook Kim, Ki-Byoung Kim, Hyoung-Joo Kim, "VIRON : An Annotation-Based Video Information Retrieval System", Seoul National Univ., 1995.
- [3] Ramesh Jain, Arun Hampapur, "Metadata in Video Databases", SIGMOD RECORD. Vol. 23. No. 4. December 1994.
- [4] William I. Grosky, "Multimedia Information Systems", IEEE, 1994.
- [5] Gulrukh Ahanger, Dan Benson, and T.D.C. Little, "Video Query formulation", SPIE Vol.2420, 1995.
- [6] Arun Hampapur, Ramesh Jain and Terry E Weymouth, "Indexing in Video Databases", SPIE Vol.2420, 1995.
- [7] 이동호, 송용준, 김형주, "SCARLET : 웨이블릿 변환을 이용한 내용 기반 이미지 검색 시스템의 설계 및 구현", 한국정보과학회논문지(c), August 1997.
- [8] 김기욱, 김형주, "비디오 주석 시스템의 설계 및 구현", 한국정보과학회 논문지 26(6), June 1997.

저자 소개



홍수열

1985.2. 경상대학교 전산통계학과 졸업 이학학사
 1994.2. 숭실대학교 컴퓨터공학과 졸업 공학석사
 1985.1~1992.6 대우전자(주) 컴퓨터개발부 근무
 1992.7~1994.2 대우통신(주) S/W연구실 근무
 1994.3~현재 동주대학 컴퓨터 정보통신 계열 조교수
 관심분야: 비주얼프로그래밍, 정보보안, 컴퓨터 그래픽스, MIS