

특집논문-02-07-2-05

동영상 요약 및 검색 시스템

정진국*, 박주현*, 낭종호*, 김경수**, 하명환**, 정병희**

Full Motion Video Abstraction and Search System

Jin-Guk Jeong*, Ju-Hyun Park*, Jong-Ho Nang*, Gyung-Su Kim**, Myung-hwan Ha** and Byung-Hee Jung**

요 약

자동으로 동영상을 요약하는 알고리즘과 고급 수준 내용 정보를 자동으로 추출하는 알고리즘은 현재 많은 연구가 진행되어 왔다. 하지만 동영상 요약이라는 것은 사용자의 특성에 따라 다양하게 나타날 수 있고, 고급 수준 내용 정보를 자동으로 추출하는 알고리즘 또한 100% 정확한 것은 아니기 때문에 이를 수정할 수 있는 도구가 필요하게 된다. 본 논문에서는 자동 알고리즘 수행 및 수동으로 수정할 수 있는 기능을 제공하기 위한 동영상 요약 및 검색 시스템을 구현했다. 동영상 요약 및 검색 시스템은 크게 서버 부분과 클라이언트 부분으로 나눌 수 있는데 서버 부분은 동영상을 요약하여 작은 동영상 파일을 만드는 부분과 검색을 위해 메타 데이터를 입력할 수 있는 부분으로 구성이 된다. 메타 데이터는 유희원칙에 따라 입력할 수 있도록 하였다. 클라이언트 부분은 서버에 의해 저장된 메타 데이터들을 이용하여 검색할 수 있는 부분과 동영상을 동적으로 요약하기 위한 부분으로 구성이 된다. 동영상 요약 부분은 MPEG-7 의 Summary Description Scheme에 맞추어서 서술된다. 본 논문에서 구현된 동영상 요약 및 검색 시스템은 대용량의 멀티미디어 데이터를 사용하는 경우 유용하게 쓰일 수 있을 것이다.

Abstract

Although there have been a lot of researches on a video abstraction algorithm, the video abstraction tool is necessary to correctly pick up some important shots in a video because the important shots might vary on the person's subjectivity. This paper proposes a video abstraction and search system for making a video abstraction automatically or manually. Proposed system has two main components, server and client. The server generates the Metadata and makes a video abstraction file. The description for video abstraction file is based on MPEG-7. Since proposed system has a function to manually correct the results of each step, the author could produce a good video abstraction easily and efficiently. The proposed system could be used to build a digital video library.

I. 서 론

최근 멀티미디어 데이터 사용이 일반화되면서 VOD(Video On Demand)나 DVL(Digital Video Library)^{[13][14]}과 같은 여러 가지 응용 분야가 나오고 있다. 이런 응용 분야에서 사용자가 네트워크를 통해서 멀티미디어 데이터의 내용을 보게

되면 멀티미디어의 큰 용량 때문에 시간이 오래 걸린다는 단점이 있다. 이러한 문제점을 해결하기 위해 많이 사용되는 방법이 동영상 요약 기법^[1]과 동영상 검색 기법이다.

동영상 요약 기법^[1]과 동영상 검색 기법^{[13][14]}에 대해서는 기존에 많은 연구들이 진행이 되어 왔다. 하지만 자동 요약 생성 알고리즘이 만족할 정도 성능을 내기 위해서는 동영상에 대한 정확한 이해를 바탕으로 한 인공지능 기법을 필요로 하는데 현재 이런 기술은 아직 개발되지 않은 상태이므로 임의의 동영상에 대하여 편집자/사용자의 요구에 맞추어서 그 내용을 자동으로 요약하는 알고리즘 및 이를 수

* 서강대학교 컴퓨터학과
Dept. of Computer Science, Sogang University

** KBS 기술연구소
KBS Technical Research Institute

동으로 편집할 수 있는 저작 도구가 필요하다고 할 수 있다. 고 수준 내용 정보에 바탕을 둔 동영상 검색 기법을 위해서도 검색을 위한 내용 정보가 필요하기 때문에 내용 정보를 입력하기 위한 저작 도구가 필요하다.

본 논문에서는 동영상 요약 및 검색 시스템을 구현했다. 크게 시스템은 서버 부분과 클라이언트 부분으로 나뉘는데 서버 부분은 정적 요약, 내용 정보 입력부와 이를 위해 선행되는 인덱싱 부분, 그리고 서버와 클라이언트를 연결하기 위한 데이터베이스로 구성된다. 클라이언트 부분은 검색을 위한 부분과 동적 요약을 위한 부분으로 구성된다. 본 논문에서 구현한 동영상 요약 및 검색 시스템은 동영상 데이터를 효율적으로 관리/검색할 수 있는 기능 및 동영상을 원하는 목적에 맞추어서 손쉽게 편집/요약할 수 있는 기능을 제공할 수 있기 때문에 동영상에 대한 편집/관리 업무에 대한 시간적, 경제적 효율성을 극대화 할 수 있을 것이다. 또한 사용자에게도 동영상에 대한 요약 기능을 제공함으로써 단시간 내에 동영상의 내용을 알아볼 수 있어 다양한 요구를 만족시키는 기능을 제공해줄 수 있을 것이다.

II. 시스템 구조

그림 1은 전체 시스템 구조를 나타내는 그림이다. 그림에서 보면 알 수 있듯이 크게 시스템은 서버측과 클라이언트측으로 구성이 된다. 서버측에서 하는 작업은 동영상 등록, 인덱싱, 요약, 동영상 정보 등록 등이고 클라이언트측에서는 서버측의 데이터베이스를 통해 원하는 정보를 질의하고 결과를 받는 작업을 한다.

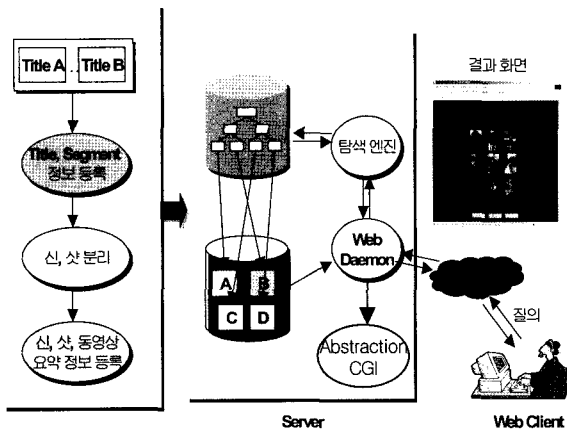


그림 1. 동영상 인덱싱 및 요약용 저작 도구 구조
Fig. 1. The structure of the video abstraction and search system

동영상을 등록하는 데 있어 가능한 파일 형식은 스트리밍이 가능한 ASF(Advanced Streaming Format) 파일 형식이다. 동영상 등록이 끝나게 된 후 행해지는 작업은 인덱싱 작업인데 인덱싱 작업이란 하나의 동영상을 샷/신으로 나누는 작업을 일컫는다. 동영상 내용 정보를 입력하는 과정이나 동영상을 요약하는 과정의 기본 단위가 되는 것이 샷 또는 신이기 때문에 이러한 작업이 필요하게 되는 것이다. 인덱싱 작업은 다시 자동 인덱싱 과정과 인덱싱 수정 과정으로 나뉘는데 자동 인덱싱 과정이라는 것은 알고리즘을 통해 동영상의 샷/신을 자동으로 찾는 과정을 말한다. 하지만 이러한 알고리즘이 100% 정확하게 모든 샷/신을 찾는 것은 아니기 때문에 샷/신을 수동으로 수정하기 위한 작업이 필요하다. 인덱싱 과정이 모두 끝나게 되면 동영상 정보 등록 과정과 동영상 요약 과정이 가능하게 된다. 이 두 과정이 끝나게 되면 서버 측의 데이터베이스에 필요한 정보 즉, 예를 들면 "A라는 사람이 B라는 장소에 위치한 신은 어떤 신인가?", "A 타이틀의 요약 파일 이름은 B이다." 등과 같은 정보들이 저장되어 데이터베이스 구축이 끝나게 되는 것이다.

클라이언트는 웹과 더불어 작동한다. 즉, 클라이언트가 어떤 장면을 원한다고 질의를 하게 되면 그 질의는 웹 서버를 통해 서버로 연결이 되고 서버 측에서 질의를 해석하여 그 결과를 웹 서버를 통해 클라이언트로 보내게 되는 것이다. 크게 질의는 두 가지로 분류할 수 있는데 하나는 검색에 관한 질의이고 다른 하나는 요약에 관한 질의이다. 검색에 관한 질의는 어떤 동영상을 원한다는 형식의 질의문으로 앞에서 설명했던 서버측 데이터베이스를 검색하여 원하는 동영상을 찾게 된다. 요약에 관한 질의는 하나의 동영상에 대해 어떤 형태의 요약을 원한다는 형식의 질의로 사용자의 요구에 따라 다양한 형식의 요약이 가능하도록 요약 CGI(Common Gateway Interface)를 통해 사용자의 입력을 받아들이도록 하였다.

III. 자동 인덱싱 알고리즘

동영상의 요약 및 검색 과정을 위해서는 신과 샷으로 동영상을 나눌 수 있는 인덱싱 과정이 필수적으로 첨가가 되어야 한다. 본 논문에서는 이와 같은 과정을 위해 자동으로 동영상을 신과 샷으로 나누어주는 알고리즘을 이용하였는데 이 장에서는 이러한 자동 인덱싱 과정을 설명하도록 하겠다.

1. 샷 경계 자동 인덱싱 알고리즘

그림 2는 본 논문에서 이용한 ASF 에서의 샷 경계 검출 방법을 나타내는 그림이다. 본 논문의 시스템은 기존의 시스템과는 다르게 ASF 파일 형식의 특징을 이용하여 보다 빠른 샷 경계 검출 방법을 구현하였다. 본 논문에서는 샷 경계를 판단하는 기준으로 인접한 프레임간의 밝기차 (Luminance Difference)를 이용하였다. 우선 인접한 두 프레임의 밝기 정보를 추출하고 히스토그램을 구한다. 본 논문에서는 히스토그램의 빈의 개수를 64개로 정하고 실험을 하였다. 히스토그램을 비교하는데 이용한 방법은 Chi-Square 방법^[11]을 이용하였다. 실험 결과 본 논문의 알고리즘은 Recall, Precision 모두 90% 정도 나오는 것을 알 수 있었다.

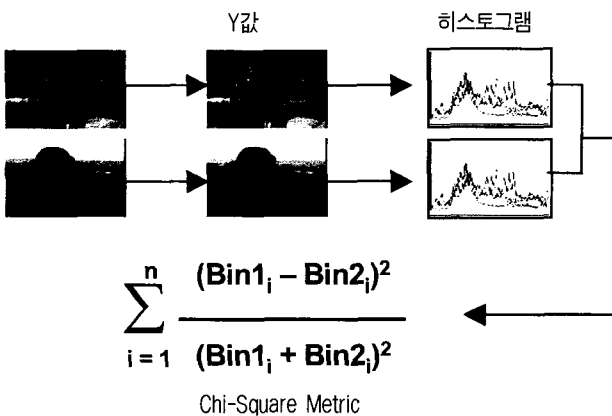


그림 2. 샷 경계 자동 인덱싱 알고리즘
Fig. 2. Shot change detection algorithm

하지만 위의 방법을 모든 프레임에 적용하면 너무 시간이 오래 걸린다는 단점이 있었다. 모든 프레임을 디코딩하는 작업이 선행되어야 하고, 각 프레임마다 빈의 개수에 따라 히스토그램을 구해야 하고 각 히스토그램의 차이를 계산해야 하기 때문이다. 29분 45초의 동영상 데이터에 대해 샷 경계 자동 인덱싱 알고리즘을 적용한 결과 약 9분 정도의 수행 시간이 걸렸다.

ASF 파일 형식의 동영상에서는 KeyFrame이라고 명명된 프레임이 중간에 있고, 인코딩을 하는 경우 KeyFrame을 선택할 수 있도록 되어 있다. 이러한 KeyFrame을 이용하면 디코딩하는 시간을 줄일 수 있다. <표 1>은 Format SDK^[12]를 이용하여 29분 45초의 동영상 데이터를 디코딩하는데 걸리는 시간을 나타내는 표이다. Format SDK는

Microsoft사에서 Windows Media 형식의 파일을 읽고 쓰고, 편집, 전송하는 데 이용하기 위해서 만든 SDK이다. 표에서 각 행의 데이터들은 해당 간격으로 프레임을 건너 뛰고 추

표 1. 프레임 추출 시간 비교
Table. 1. Comparison of decoding time

	실행시간	전체 프레임 개수	프레임당 추출 시간
KeyFrame만	1분 15초	710프레임	0.1초
180프레임당	1분 26초	297프레임	0.28초
120프레임당	2분	445프레임	0.26초
90프레임당	2분 13초	593프레임	0.22초
60프레임당	3분	890프레임	0.20초
30프레임당	7분	1780프레임	0.24초

출하는 경우의 실행 시간 정보를 나타낸다. 표를 보면 알 수 있듯이 KeyFrame-ASF에서는 참조 프레임인 KeyFrame을 비트율에 따라 여러 개의 프레임 사이에 위치시키는데 보통 샷이 변하는 위치에 KeyFrame이 위치되어진다. -만 추출하는 경우가 다른 경우보다 프레임당 디코딩 시간이 빠른 것을 알 수 있다. 또한 전체 프레임 개수와 프레임당 추출 시간을 살펴 보면 전체 프레임 개수가 KeyFrame과 가까운 경우가 다른 경우보다 프레임당 추출 시간이 빠른 것을 알 수 있다. 이와 같은 실험 결과를 통해 Format SDK를 이용해 프레임을 추출하는 경우에는 KeyFrame을 중심으로 해당 프레임을 찾아가기 때문에 프레임당 추출 시간이 KeyFrame만 추출하는 경우가 가장 빠르게 나온다는 것을 알 수 있었다. 그러므로 본 시스템에서는 전체 프레임에 알고리즘을 적용하지 않고 KeyFrame에만 알고리즘을 적용하여 성능을 높이는 방안을 고려하였다. 실험 결과 29분 45초의 같은 동영상 데이터에 대해 1분 15초 정도의 수행 시간에 비슷한 성능을 보였다.

2. 신 경계 자동 인덱싱 알고리즘

샷은 인덱싱의 기본적인 단위로서 그 경계를 검출하기는 쉽지만 하나의 비디오 데이터에 속하는 샷의 개수는 너무나 방대하며 하나의 샷이 어떠한 의미를 가진다거나 사용자의 질의에 대한 답이 되기엔 해당 샷이 가진 정보가 너무 부족하게 된다. 이러한 경우 이용될 수 있는 단위가 신이다. 신은 하나의 의미적인 단위로 인식되어 그 적절한 대상이 된다. 기존에 나온 신 경계 자동 인덱싱은 크게 비디오 데이터를 이용한 방법^[5]과 오디오 데이터를 이용한 방법^[6], 비디오, 오디오 데이터를 같이 이용하는 방법^[7]으로 나눌 수 있다. 본 논문에서는 첫 번째 방법인 비디오 데이

터를 이용한 방법을 이용하였다.

그림 3은 신 경계 자동 인덱싱 알고리즘을 나타내는 그림이다. 전 프레임과 현재 프레임 두 개의 프레임이 존재할 때 현재 프레임과 전 프레임의 같은 위치에 존재하는 픽셀

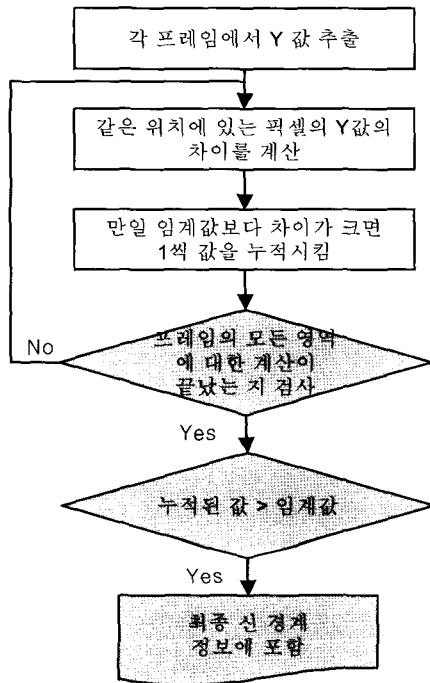


그림 3. 신 경계 자동 인덱싱 알고리즘
Fig. 3. Scene change detection algorithm

의 Y값을 비교하여 임계값을 넘는 픽셀이 많은 경우 신이 변했다고 결정을 하게 된다. 보통 같은 신안에서는 배경이 비슷하기 때문에 이러한 알고리즘을 이용하면 배경이 변하는 신의 경계를 찾을 수 있게 된다.

IV. 동영상 자동 요약 알고리즘

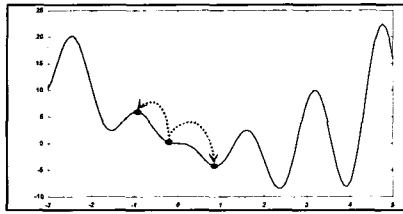
동영상 요약은 본래의 동영상이 가지는 핵심 메시지를 유지한 상태에서 축약된 정보를 가지는 정지화면이나 동영상을 만드는 것을 말한다. 동영상 요약 알고리즘은 기존에 여러 가지 방향으로 진행이 되어 왔다. 예를 들면 비슷한 신들을 그룹화하여 각 그룹에서 신들을 추출하는 방법^[1], 움직임이 많은 샷들을 추출하는 방법^[8], 시간대별로 샷들을 추출하는 방법^[9]등을 들 수 있는데 이러한 기존의 방법들은 사용자의 다양한 요구를 충족시켜줄 수 없다는 단점을

가지고 있다.

본 논문에서는 목적 함수를 이용한 동영상 요약 알고리즘^[10]을 시스템에 적용시켰다. 목적 함수를 이용한 동영상 요약 알고리즘이라는 것은 동영상 요약을 이루는 샷들이 가져야 하는 바람직한 특성들을 목적함수로 표현하고, 이런 목적함수를 이용하여 사용자가 정하는 가중치에 따라 가장 높은 샷들의 집합을 찾아내는 알고리즘이다. 목적함수로 사용하는 것은 Well Distributed, Well Fitting, Event, Concise, Not too short, Shot Exclusion, Non Bias 7가지이다. Well Distributed는 전체에 걸쳐 고르게 샷이 선택되어야 한다는 것이고 Well Fitting은 사용자가 원하는 시간에 맞도록 요약 파일이 만들어져야 한다는 것이다. Event는 움직임이 많은 샷이 선택이 되어야 한다는 것이고 Concise는 선택되는 샷들 사이에 중복이 없어야 한다는 것이며 Not too short는 너무 짧은 샷이 선택되어서는 안되게 하는 목적 함수이다. Shot Exclusion은 결론 부분은 보여주지 말아야 한다는 목적 함수이고 마지막으로 Non Bias는 하나의 샷이 전체 요약 파일의 대부분을 차지해서는 안되게 하는 목적 함수이다. 이러한 7가지 목적 함수에서 사용자가 원하는 목적 함수에 큰 가중치를 두어 사용자가 원하는 요약 파일을 만들게 되는 것이다. 사용자가 원하는 목적 함수에 가중치를 주면 각 목적 함수의 값과 가중치가 곱해져서 정규화된 후에 합해지기 때문에 사용자의 의도가 반영되게 되는 것이다. 기존에 연구된 요약 알고리즘들은 목적 함수에 주어지는 가중치를 이용하여 표현 가능하다. 예를 들어 시간 축을 바탕으로 하는 요약은 Well Distributed, Well Fitting에 가중치를 높게 주어 요약을 하면 된다. 이 알고리즘의 장점은 확장과 주관적 요약이 가능하다는 데 있다. 여기서 확장은 요약 길이를 변화시킴에 따라 목적함수가 변화될 수 있다는 장점이고, 주관은 이러한 목적함수에 대한 가중치를 변경함으로써 요약자의 주관을 반영시킬 수 있는 장점을 말한다.

하지만 위에서 설명한 목적 함수를 이용하는 알고리즘에서는 목적 함수를 최대화하는 집합을 찾아내는 데 시간이 오래 걸린다는 단점이 있다. 즉, n 개의 샷으로 구성된 동영상을 위해서는 각 샷에 대해 요약 파일에 첨가를 시킬 것인지 그렇지 않을 것인지를 결정하기 위해 2n번의 계산이 필요하기 때문에 전체 알고리즘을 수행하는 데 시간이 오래 걸리는 데 이를 해결하기 위해 Simulated Annealing (SA) 알고리즘^[15]을 이용한다. 그림 4는 SA 알고리즘을 설명하는 그림이다. SA 알고리즘은 기본적으로 국소 탐색 방법을 개선한 방법으로 조합 최적화 문제에 많이 쓰이는

알고리즘이다. 이 알고리즘을 이용하여 최소값의 근사값을 빠른 시간 내에 찾아낼 수 있다. 이 알고리즘을 좀 더 자세히 설명하면 먼저 T라는 초기 온도를 정의하는 데 이는 종료하는 조건을 위한 것이다. 우선 하나의 shot집합을 임의로 뽑고, 이렇게 뽑은 shot집합의 목적함수와 가중치들의 합(e1)을 구한다. 또 다시 하나의 shot집합을 뽑고 목적함수



1. 임의로 혹은 다른 방법-보통 Heuristic-을 이용하여 C_1 이라는 초기값을 설정한다. 이 값은 연결된 $E_i = F(C_i)$ 이라는 함수를 갖게됨.
2. 초기 온도(temperature) T가 정의됨
3. 임의의 변화에 의해 새로운 값 C_{i+1} 로 새롭게 설정됨
4. 만일 $E_{i+1} \leq E_i$ 라면 변화는 항상 받아들여짐
5. 만일 $E_{i+1} > E_i$ 라면 변화는 확률을 가지고 받아들여짐
6. 온도(temperature) T에 설정된 수(무한번 반복하기 위해서는 1로 설정)만큼 곱한 후에 2로 돌아간다. 이러한 과정을 T가 조건보다 작아질 때까지 반복함.

그림 4. SA 알고리즘
Fig. 4. SA Algorithm

와 가중치의 합(e2)을 구한다. 이렇게 다시 뽑은 shot들의 집합의 값(e2)가 e1보다 크게 되면 이 값(e2)와 이 집합을 현재의 shot집합으로 결정한다. e2가 작게 되는 경우에는 e1, e2와 T로 Boltzman 확률(p) $p = e^{(e1 - e2)/kT}$ - 를 구하고 이를 만족하는지(Condition())를 구하여 만족하게 되면 e2와 현재 shot집합을 채택하고, 아니면 T의 값에 틀 곱하여 줄이게 된다. 이렇게 하여 T가 보다 작게 되면 현재의 shot집합을 채택함으로써 종료하게 된다.

V. 동영상 인덱싱 및 요약용 저작 도구

1. 동영상 인덱싱 및 요약용 저작 도구

그림 5는 동영상 인덱싱 및 요약용 저작 도구 구조를 나타내는 그림이다. 동영상 인덱싱 및 요약용 저작 도구는 동영상을 요약하는데 있어서 기본 단계인 샷/신 단위 인덱싱과 고급 수준 내용 정보를 입력 할 수 있으며, 그 결과를 토대로 요약된 동영상을 자동 또는 수동으로 만들어내는

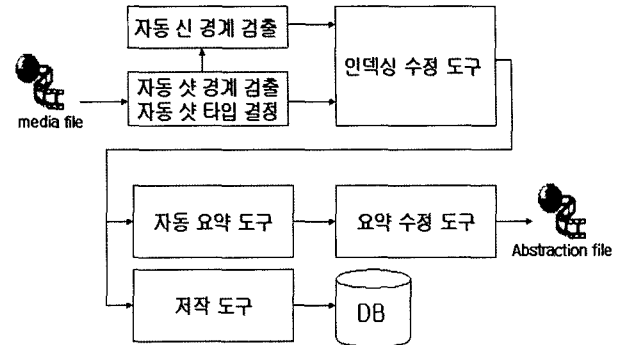


그림 5. 동영상 인덱싱 및 요약용 저작 도구 구조
Fig. 5. The structure of the indexing tool and the authoring tool

도구이다. 전체 시스템은 크게 여섯 가지 도구 자동 샷 경계 검출/자동 샷 타입 결정 도구, 자동 신 경계 검출 도구, 인덱싱 수정 도구, 자동 요약 도구, 요약 수정 도구, 저작 도구 - 의 조합으로 이루어진다.

자동 샷 경계 검출/자동 샷 타입 결정 도구는 3.1 장에서 설명했던 알고리즘을 이용하여 자동으로 샷 경계를 검출해주고 동영상의 장르가 뉴스나 축구일 경우 샷의 타입을 결정해 주는 도구이다. 또한 이 과정에서는 요약에 필요한 저수준 내용 정보 움직임 에너지, Optical Flow 값, 칼라 히스토그램 등을 추출하게 된다. 자동 신 경계 검출 도구는 3.2 장에서 설명했던 알고리즘을 이용하여 자동으로 신 경계를 검출해주는 도구이다.

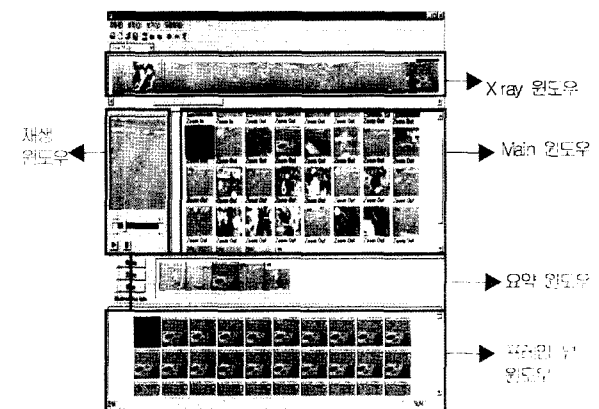


그림 6. 사용자 인터페이스
Fig. 6. User Interface

자동 알고리즘들이 수행이 된 이후에는 이러한 자동 알고리즘들이 100% 정확한 것은 아니기 때문에 수동으로 이를 수정하기 위한 도구가 필요하다. 그림 6은 이러한 도구

들을 위한 전체 인터페이스를 보여주는 그림이다. 본 시스템 인터페이스의 특징은 하나의 작업 환경에서 모든 기능이 이루어질 수 있도록 구현되었다는 점이다. 그림에서 상단에 있는 윈도우는 X-Ray 윈도우로 각 프레임의 대각선을 60픽셀씩 추출하여 배치함으로써 동영상의 전체 구조를 파악할 수 있도록 해주는 윈도우이다. 좌측에 있는 재생 윈도우는 선택된 샷을 재생해주는 윈도우이고, 중앙에 있는 메인 윈도우는 전체 샷의 대표 프레임들을 보여주며 샷/신경계 수정, 샷 타입 수정 등의 작업을 하는 윈도우이다. 메인 윈도우 아래 위치하는 요약 윈도우는 요약 동영상 내의 샷들의 모임을 대표 프레임과 길이 정보를 통해 보여주고 편집할 수 있도록 해주는 윈도우이다. 마지막으로 하단에 위치한 프레임 뷰 윈도우는 선택된 샷의 모든 프레임들을 볼 필요가 있는 경우 모든 프레임을 실시간으로 디코딩 하여 보여주는 기능을 가지고 있다.

인텍싱 수정 도구는 그림 6의 메인 윈도우와 프레임 뷰 윈도우를 이용하는 것이다. 즉, 메인 윈도우를 이용하여 샷의 병합, 신의 병합/나누기가 가능하고, 프레임 뷰 윈도우를 통해 샷을 나누는 작업이 가능한 것이다. 그림 7은 자동 요약 도구를 위한 인터페이스를 보여주는 그림이다. 이 인터페이스를 통해 사용자는 필요한 가중치와 매개변수를 설정하여 원하는 요약 파일을 만들 수 있다. 하지만 이러한 자동 알고리즘을 통해 만든 요약 파일이 사용자의 요구를 완전하게 만족시킬 수 없기 때문에 수정 도구가 필요하게

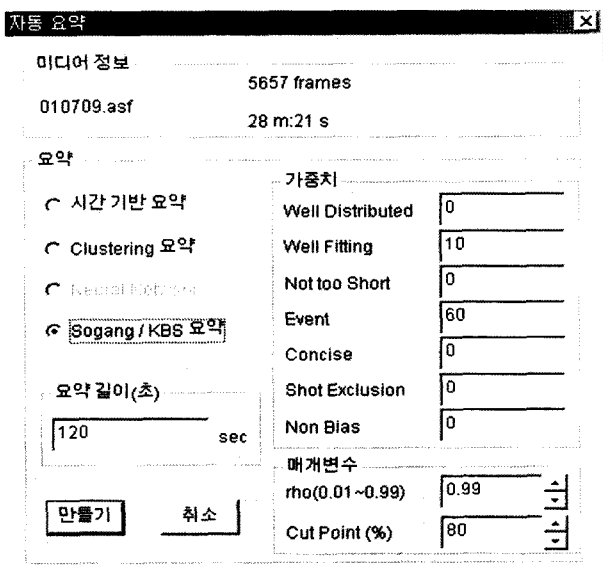


그림 7. 자동 요약 인터페이스
Fig. 7. User Interface for automatic video abstraction

되는데 요약 수정 도구는 요약 윈도우와 메인 윈도우를 이용하게 된다. 전체 요약 파일에서 원하지 않는 샷을 제거하는 것은 요약 윈도우를 통해 작업할 수 있고 새로운 샷을 요약 파일에 첨가하는 일은 메인 윈도우에서 샷을 드래그하여 요약 윈도우에다 넣어주면 된다.

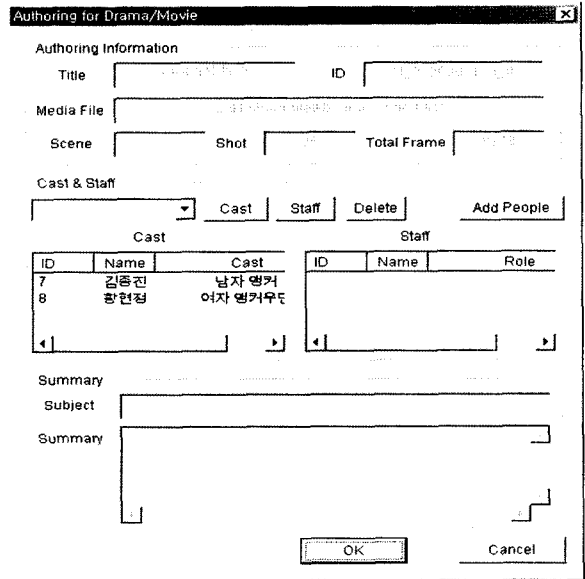


그림 8. 타이틀 단위 내용 정보 입력 인터페이스
Fig. 8. User Interface for recording the title

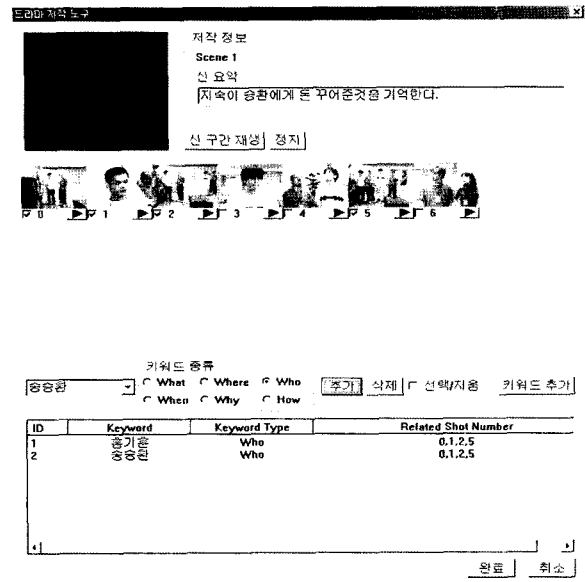


그림 9. 신 단위 내용 정보 입력 인터페이스
Fig. 9. User Interface for recording contents of scene

내용 정보를 입력하기 위한 저작 도구는 그림 6의 사용자 인터페이스에서 모드를 바꾸어서 수행이 된다. 그림 8은 타이틀 단위 내용 정보 입력 인터페이스를 나타내고 있고 그림 9는 신 단위 내용 정보 입력 인터페이스를 나타내는 그림이다. 타이틀 단위 내용 정보 입력에서는 타이틀을 만든 스태프나 배우 등의 정보를 입력하게 되고 신 단위 내용 정보 입력에서는 각 신의 내용을 요약할 수 있는 정보를 육하 원칙에 따라서 입력하게 된다.

위와 같은 인터페이스와 이를 이용하는 여섯 가지 도구들을 통해서 사용자는 원하는 내용들을 샷/신 단위로 데이터베이스에 저장하여 관리할 수 있고 원하는 요약 파일을 생성할 수 있게 되는 것이다.

2. 데이터베이스 설계

많은 비디오 데이터와 고급 수준 내용 정보를 효율적으로 관리하고 사용자에게 브라우징이나 검색 등의 기능을 제공하기 위해서는 데이터베이스를 이용해야 한다. 본 논문에서는 이를 위해 관계형 데이터베이스로 데이터베이스를 구현하였다. 기본적으로 여러 장르를 하나의 데이터베이스로 관리하는 것이 특징이다. 즉, 뉴스나 축구 등 각각의 특성을 살릴 수 있으면서도 중복되는 부분은 하나로 합치고 있다. 이러한 데이터베이스를 통하면 다른 장르로도 쉽게 확장할 수 있다.

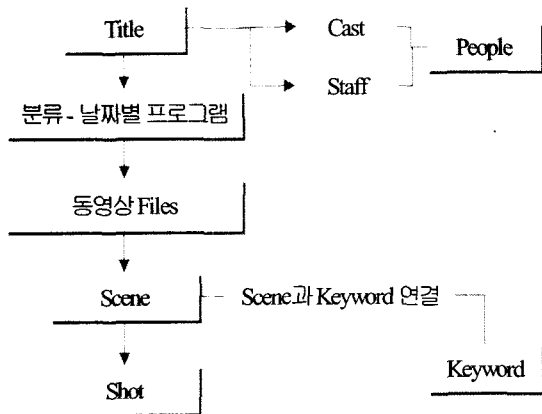
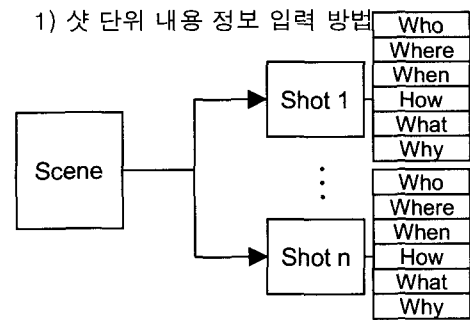


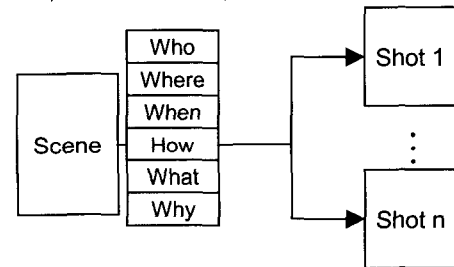
그림 10. 테이블 관계도
Fig. 10. Relational Table

그림 10은 본 시스템의 데이터베이스 테이블 관계도를 나타내는 그림이다. Title은 최상위 레벨의 고급 수준 내용 정보로서 타이틀의 종류와 제작사, 국가에 대한 정보가 들

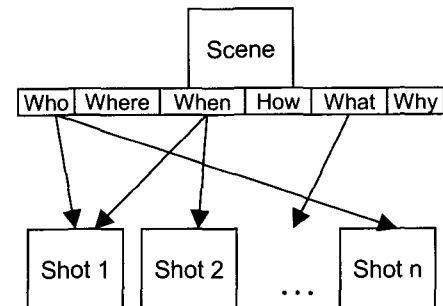
어가며 유일한 아이디를 부여 받는다. Title은 그 종류에 따라서 각각 뉴스, 드라마, 영화, 축구로 나뉘어 진다. 영화의 개봉날짜와 드라마의 방영날짜와 같이 비슷한 의미를 갖는 내용들이 있지만 확장성과 좀 더 정확한 의미 부여를 위해서 각각의 테이블들은 통합되지 않고 독립된 테이블로 두었다. 따라서 Title의 종류가 정해지면 최하위 계층인 샷까지 계층적인 구조가 정해진다. 영화를 제외한 뉴스, 드라마, 축구는 주기적으로 TV에 방영되므로 날짜별로 구분될 수 있도록 하였다.



1) 샷 단위 내용 정보 입력 방법



2) 신 단위 내용 정보 입력 방법



3) 혼합된 내용 정보 입력 방법
그림 11 내용 정보 입력 방법
Fig. 11. Input method of contents information

고급 수준 내용 정보 데이터베이스는 육하 원칙에 따라 구분하였다. 고급 수준 내용 정보를 입력하는 방법은 크게 세 가지 방법이 있을 수 있다. 그림 11은 세 가지 방법들

비교하는 그림이다. 첫 번째는 샷 단위로 고급 수준 내용 정보를 입력하는 것이고, 두 번째는 신 단위로 고급 수준 내용 정보를 입력하는 것이다. 마지막 방법은 첫 번째 방법과 두 번째 방법을 혼합하여 필요한 샷에 대해서만 고급 수준 내용 정보를 입력하는 것이다. 즉, 'Who'라는 키워드에 'A'라는 사람이 입력이 되었고 신내에 있는 샷 중에 한 두 샷에만 'A'라는 사람이 나온다면 이 샷들만 키워드와 연결을 시키는 것이다. 첫 번째 방법은 샷 단위의 검색이 가능하다는 장점이 있지만 중복되는 내용이 너무 많아지고, 한 비디오 내에 너무 많은 샷이 존재하기 때문에 내용 정보를 입력하는 데 너무 많은 시간이 걸린다는 단점이 있다. 이에 반해 두 번째 방법은 첫 번째 방법에 비해 중복되는 내용이 없고 내용 정보를 입력하는 데 많은 시간이 걸리지는 않지만 샷 단위의 검색이 불가능하다는 단점이 있다. 세 번째 방법은 두 번째 방법보다는 내용 정보를 입력하는 데 시간이 오래 걸리기는 하지만 첫 번째 방법보다는 빠르게 내용 정보를 입력할 수 있다. 또한 중복되는 내용 정보도 줄일 수 있고 샷 단위로 내용 정보들이 입력이 되기 때문에 샷 단위의 검색도 가능하게 된다. 본 시스템에서는 이러한 이유로 세 번째 방법을 이용한다.

VI. 웹에서의 검색 기능 설계

웹 서비스를 위해서 사용한 웹 서버는 인터넷 정보 서비스(IIS)이며, 인터넷 정보 서비스에서 작동하는 ASP를 이용하여 웹 프로그래밍을 하였다.

1. 제공하는 주요 기능

제공하는 기능은 크게 검색에 관한 기능과 요약에 관한 기능으로 나눌 수 있다. 기존의 시스템들 대부분이 검색이나 요약 중 하나의 기능만 제공하는 데 반해 본 논문의 시스템은 검색, 요약에 관한 기능을 같이 제공하여 사용자가 쉽게 자신의 의도대로 사용할 수 있도록 하였다. 그림 12는 뉴스에서의 검색 조건을 나타내는 그림이다. 기본적으로 육하원칙에 의해 질의를 할 수 있고, 기사 타입과 정치, 경제 등의 섹션에 관한 질의가 가능하다. 드라마 또한 뉴스와 비슷하게 육하원칙에 따른 질의가 가능하다. 하지만 드라마에서는 섹션이나 기사 타입이 존재하지 않고 타이틀이 존재하기 때문에 섹션, 기사 타입에 대한 질의는 불가능하고 타이틀에 대한 질의가 가능하다. 스포츠에서는 육하원칙이 의

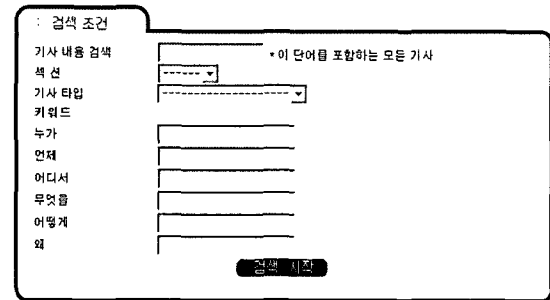


그림 12. 뉴스에서의 검색 조건
Fig. 12. Search condition of the news

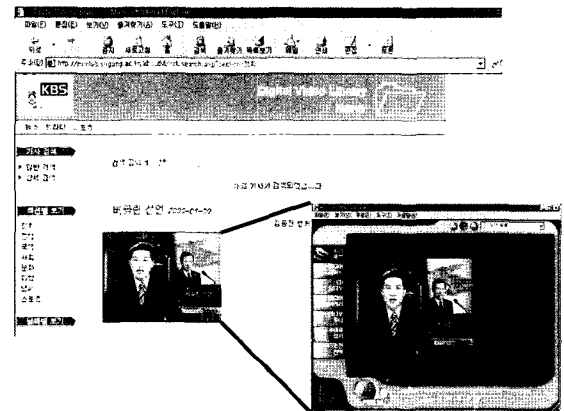


그림 13. 검색 결과
Fig. 13 Search result

미가 없으므로 대회, 리그 이름, 팀, 선수 이름 등에 의해서 질의가 가능하다. 그림 13은 질의에 대한 결과 화면을 보여주는 그림이다. 그림에서 보는 바와 같이 사용자에게는 대표 프레임과 간단한 신의 내용이 보여지고 대표 프레임을 클릭하면 동영상에서 해당하는 신만 실시간으로 클라이언트로 보내지고 WMP(Windows Media Player)를 통해 재생되게 된다. 이런 모든 질의가 가능한 이유는 5장에서 설명했던 도구들을 통해 서버측의 데이터베이스가 구축되었기 때문이다.

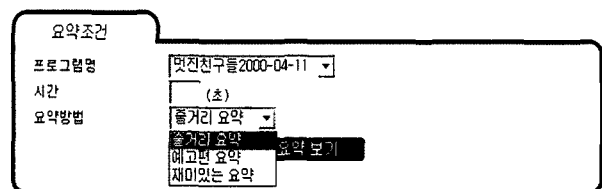


그림 14. 드라마에서의 요약 조건
Fig. 14. Search condition of the drama

그림 14는 드라마 장르 동영상의 요약 조건을 보여주

는 그림이다. 여기서 언급되는 요약이라는 것은 동적 요약을 말한다. 즉, 사용자가 웹 브라우저를 통해 자신이 원하는 요약이 어떤 요약인지를 입력하면 그에 해당하는 동영상은 실시간으로 만들어져 사용자에게 보여지는 것이다. 현재는 크게 세 가지로 구분이 된다. '줄거리 요약', '예고편 요약', '재미있는 요약'이 그것인데, 실제 내부적으로는 4장에서 언급되었던 목적 함수의 가중치를 조절하여 만들어지는 것이다. '줄거리 요약'이라는 것은 전체 동영상 내에서 샷이 고르게 선택이 되어야 하기 때문에 Well Distributed에 가중치가 크게 주어질 것이다. '예고편 요약'이라는 것은 동영상의 결론이 사용자에게 보여지면 안되기 때문에 Shot Exclusion에 가중치가 크게 주어질 것이고, '재미있는 요약'이라는 것은 Event에 가중치가 크게 주어질 것이다. 이 외에도 목적 함수의 가중치에 따라 여러 가지 요약 조건이 가능할 것이다. 본 시스템에서는 이러한 계산이나 샷의 선택이 사용자의 입력에 따라 동적으로 이루어져야 하기 때문에 CGI로 구현이 되어 웹 서버와 연동이 된다.

그림 15. 뉴스에서의 요약 조건
Fig. 15. Abstraction condition of the news

그림 16 스포츠에서의 요약 조건
Fig. 16. Abstraction condition of the sports

그림 15와 그림 16은 각각 뉴스에서의 요약 조건과 스포츠(축구)에서의 요약 조건을 나타내는 그림이다. 뉴스와 스포츠 같은 경우에는 장르 고유의 특성이 있기 때문에 드라마와는 다르게 요약 조건이 적용된다. 뉴스에서는 섹션, 기사 형태, Report 타입에 따라 요약이 된다. 섹션은 정치, 경제 등의 분야를 나타내고 기사 형태는 사건 사고, 현장 추

적 등 기획 기사의 형태를 나타낸다. Report 타입은 각 샷의 타입을 얘기하는데 예를 들면 앵커, 그래픽 등을 들 수 있다. 스포츠에서는 골과 하이라이트 두 가지 요약이 있을 수 있다. 골은 저작 도구를 통해 골이라고 입력한 샷들만 선택이 되고 하이라이트는 골 이외에 중요하다고 여겨지는 샷들도 선택이 된다. 하이라이트에서는 4장에서 설명했던 동영상 요약 알고리즘이 수행이 되는데 약간 변형된 형태로 동작을 한다. 즉, 4장에서 설명했던 동영상 요약 알고리즘에서는 'Event' 목적 함수가 움직임이 많은 샷이라고 정의가 되었는데 여기서는 실제 스포츠에서 벌어지는 이벤트들 - 재연 장면, 골 장면, 축소 장면, 확대 장면 - 로 정의를 했다. 알고리즘에서는 골 장면일 때 가장 높은 가중치를 주고 다음은 재연 장면 이전 샷, 확대 장면, 축소 장면 순으로 높은 가중치를 두어 골 장면과 재연 장면 이전 샷 위주로 요약을 만들게 하였다.

2. MPEG-7 과의 연계

본 논문에서는 MPEG-7과의 연계를 위하여 MPEG-7내의 SequentialSummary DS를 이용하였다. SequentialSummary DS는 이미지나 오디오와 동기화된 비디오 프레임들의 요약 기술하기 위해서 사용된다. SequentialSummary DS는 시간적으로 연속적인 프레임들의 집합으로 요약 동영상을 표현할 수 있으며 오디오와 비디오 파일로 요약 동영상을 표현할 수도 있다. 그림 17은 SequentialSummary DS의 예를 보여주는 그림이다. 처음의 locator는 원본 비디오의 저장 위치를 나타내며 두 번째의 locator는 요약된 비디오의 위치를 나타낸다.

```
<SequentialSummary id="SoccerSummary001">
  <SourceLocator>
    <MediaURI>file://disk/soccer001.mpg</MediaURI>
  </SourceLocator>
  <VideoSummaryLocator>
    <MediaURI>file://disk/soccer001.mpg</MediaURI>
  </VideoSummaryLocator>
  <FrameProperty> . . . </FrameProperty>
  <FrameProperty> . . . </FrameProperty>
  <FrameProperty> . . . </FrameProperty>
</SequentialSummary>
```

그림 17. SequentialSummary DS의 예
Fig. 17. A example of "SequentialSummary DS"

본 시스템에서는 이러한 SequentialSummary를 적용하기 위하여 웹에서 XML로 작성된 SequentialSummary를 볼 수 있도록 하였다. 즉, 동적 요약을 생성시킬 때마다 Sequential-

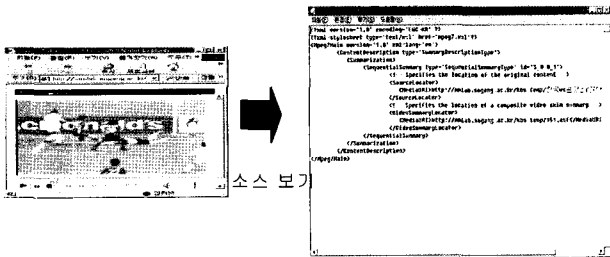


그림 18. XML 문서의 예
Fig. 18. A example of XML document

Summary DS가 하나씩 생기는 것이다. 그림 18은 동적 요약에 의해 나오는 SequentialSummary DS를 보여주는 그림이다. 본 시스템에서는 웹에서 MPEG-7에 따르는 문서를 보여주기 위해서 XSL을 이용한다. 즉, MPEG-7은 XML을 근간으로 형성된 문서이므로, 웹에서의 서비스는 XML의 스타일 문서인 XSL을 사용하여 사용자에게 제공된다. 그림 19는 XSL의 동작 원리를 나타내는 그림이다. XML문서가 갖고 있는 고유의 패턴을 인식하고 XSL이 갖고 있는 패턴에 해당하는 규칙을 적용하여 일반적인 웹 문서처럼 사용자에게 제공할 수 있도록 하는 것이다.

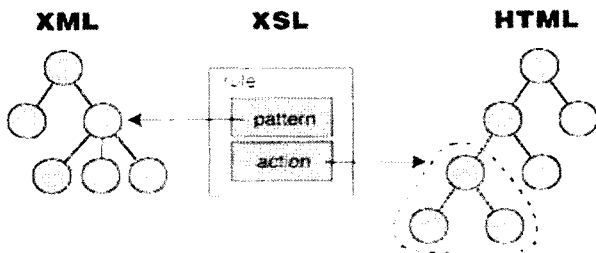


그림 19. XSL의 동작 원리
Fig. 19. The action principle of XSL

Ⅶ. 결 론

최근 멀티미디어 데이터 사용이 일반화되면서 VOD, DVL과 같은 많은 응용 분야가 발전하고 있다. 이런 응용 분야에서 문제로 인식되는 것이 사용자가 네트워크를 통해서 멀티미디어 데이터의 내용을 보는 데 큰 용량으로 인해 시간이 오래 걸린다는 점이다. 이러한 문제점을 해결하기 위해 많이 이용되는 방법이 동영상 자동 요약 기법과 동영상 검색 기법이다. 동영상 자동 요약 기법은 동영상의 내용 정보를 유지하는 상태에서 작은 동영상 혹은 대표 프레임들의 집합을 만드는 기법을 말하고 동영상 검색 기법은 사

용자가 여러 개의 동영상 파일 중에서 원하는 동영상을 찾을 수 있도록 해주는 기법을 말한다. 하지만 이러한 기법들이 아직 사용자가 만족할 만큼의 수준이 아니기 때문에 사람이 수동으로 수정할 수 있는 도구가 필요하게 된다.

본 논문에서는 동영상 요약 및 검색 시스템을 구현했다. 기존의 여러 시스템과는 다르게 검색과 요약에 관한 기능을 한 시스템 내에서 가능하도록 하였고, 현재 인터넷에서 많이 쓰이는 ASF 파일 형식을 이용함으로써 현재 실제로 사용될 수 있는 시스템을 구성하였으며 ASF 파일 형식의 특징을 이용하여 샷 자동 경계 검출 방법의 속도를 크게 향상시켰다.

시스템의 구조는 다음과 같다. 크게 서버 부분과 클라이언트 부분으로 나뉘는데 서버 부분은 사용자가 원하는 동영상을 검색하고 미리 만들어진 동영상 요약 파일을 보기 위하여 필요한 데이터베이스를 구축하기 위한 도구들을 말한다. 클라이언트 부분은 이러한 데이터베이스 기반으로 동영상을 검색하는 기능과 동적으로 사용자의 요구에 의해 동영상 요약 파일을 만드는 기능을 가지고 있다.

본 논문에서 구현한 시스템을 이용하여 인터넷 서비스에서 동영상에 대한 검색/요약 기능을 제공함으로써 적절하게 사용자의 요구를 만족시키는 기능을 제공해 줄 수 있다. 또한 본 논문에서 개발되는 다양한 기술들을 이용하여 보다 효율적으로 멀티미디어 데이터를 저장, 관리할 33수 있을 것이다. 앞으로는 각 장르에 알맞은 인덱싱 알고리즘, 요약 알고리즘을 개발하여 보다 견고한 동영상 요약 및 검색 시스템을 구축하도록 할 것이다.

참 고 문 헌

- [1] Silvia Pfeiffer, Rainer Lienhart, Stephan Fischer and Wolfgang Effelsberg, "Abstracting Digital Movies Automatically," *Journal of Visual Communication and Image*, Vol. 7, No. 4, pp.345-353, 1996.
- [2] H. Jiang, A. Helal, A. Elmagarmid and A. Joshi, "Scene change detection techniques for video database systems," *Proceedings of ACM Multimedia Systems*, 1998.
- [3] S.C. Pei, "Efficient MPEG Compressed Video Analysis Using Macroblock Type Information," *IEEE Transactions on Multimedia*, Vol. 1, No.4, 1999.
- [4] 이화순, "MPEG 비디오 스트림의 칼라 히스토그램 정보와 매크로블록 정보를 이용한 새로운 샷 경계 검출 방법," *서강대학교 석사 학위 논문*, 2000.
- [5] 이숙경, "MPEG 비디오 스트림에서 줄거리 특성에 기초한 신 경계 검출 방법," *서강대학교 석사 학위 논문*, 1998.
- [6] 김재홍, "MPEG 시스템 스트림 상에서 오디오 정보를 이용한 신경계

- 검출 방법”, 서강대학교 석사 학위 논문, 1999.
- [7] 강찬미, “MPEG 스트림에서의 비디오 및 오디오 정보를 이용한 신 경계 검출 방법”, 서강대학교 석사 학위 논문, 2000.
- [8] J. Nam and A. H. Tewfik, “Video Abstract of Video,” *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP '99)*, pp. 117-122, 1999.
- [9] Rainer Lienhart, “Dynamic Video Summarization of Home Video,” *Proceedings of Storage and Retrieval for Media Databases 2000* pp. 378-389, 2000.
- [10] 홍승욱, “MC-DCT 영역에서의 비디오 요약 시스템”, 서강대학교 박사 학위 논문, 2001.
- [11] I. K. Sethi and Nilesh V. Patel, “A Statistical Approach to Scene Change Detection,” *Proceedings of SPIE on Storage and Retrieval for Image and Video Databases*, pp 329-339, 1995.
- [12] Microsoft Inc, Windows Media SDK 홈, <http://www.microsoft.com/korea/msdn/workshop/imedia/windowsmedia/sdk/wmsdk.asp>, 2001.
- [13] A. G. Hauptmann and M. J. Witbrock, “Story Segmentation and Detection of Commercials in Broadcast News Video,” *Proceedings of Advances in Digital Libraries*, 1998.
- [14] Carnegie Mellon University, Informedia Digital Video Library at CMU, <http://www.informedia.cs.cmu.edu>.
- [15] R.H.J.M. Ottern and L.P.P.P van Ginneken, *The Annealing Algorithm*, Kluwer Academic Publishers, 1989.

저 자 소 개



정진국

- 1998년 : 서강대 전자계산학과 졸업
- 2000년 : 서강대 컴퓨터학과 석사
- 현재 : 서강대 컴퓨터학과 박사과정 재학중
- 주관심분야 : Digital Video Library, 미디어 아카이브



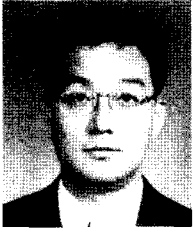
박주현

- 1999년 : 서강대 전자계산학과 졸업
- 2002년 : 서강대 컴퓨터학과 석사
- 현재 : 서강대 컴퓨터학과 박사과정 재학중
- 주관심분야 : Digital Video Library, 미디어 아카이브



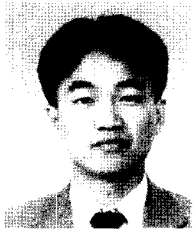
남종호

- 1986년 : 서강대 전자계산학과 졸업
- 1988년 : 한국과학기술원 석사
- 1992년 : 한국과학기술원 박사
- 1992년~1993년 : Fujitsu 연구소 연구원
- 1993년~현재 : 서강대학교 컴퓨터학과 부교수
- 주관심분야 : Digital Video Library, 미디어 아카이브



김 경 수

- 1983년 : 서울대학교 제어계측공학과 졸업
- 1985년 : 서울대학교 제어계측공학과 석사
- 1985년~현재 : KBS 기술연구소 차장
- 주관심분야 : 멀티미디어 방송 제작 시스템, 스트리밍 미디어 및 서비스



하 명 환

- 1993년 : 경북대학교 전자공학과 졸업
- 1995년 : 한국과학기술원 전기 및 전자공학과 석사
- 1995년~현재 : KBS 기술연구소 연구원
- 주관심분야 : 멀티미디어 방송 제작 시스템, 비디오 인덱싱



정 병 희

- 1994년 : 이화여자대학교 전자계산학과 졸업
- 1996년 : 한국과학기술원 전산학과 석사
- 1996년~현재 : KBS 기술연구소 연구원
- 2000년~현재 : 한국과학기술원 전산학과 박사과정 재학중
- 주관심분야 : 멀티미디어 방송 시스템, 미디어 아카이브