

碩 士 學 位 論 文

MPEG-2 비디오 데이터에서 내용기반
검색에 관한 연구



濟州大學校 産業大學院

電子電氣工學科

朴 宰 必

2000年 6月

碩 士 學 位 論 文

MPEG-2 비디오 데이터에서 내용기반
검색에 관한 연구

指導教授 金 壯 亨

이 論文을 工學碩士學位 論文으로 提出함.

2000年 6月 日

 濟州大學校 産業大學院
제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY
電子電氣工學科(컴퓨터 工學:)

朴 宰 必

朴宰必의 工學碩士學位 論文을 認准함.

2000年 6月 日

委員長 安 基 中

委 員 金 壯 亨

委 員 李 尙 俊



A Study on contents-base retrieval of MPEG-2 Video data

Jae-pil Park

Department of Electrical and Electronic Engineering
Graduate School of Industry
Cheju National University

Supervised by professor Jang-Hyung Kim



Summary

Recent progresses on computer communication network, multimedia information processing and communication service technology, accelerate the entrance of information society. But, WWW brought the information crisis in its demand and size on the communication network, by making it easy to develop information services, like electronic news, electronic conference, multimedia information retrieval, and so on. Due to this changes, DBMS should provide efficient ways to store and manage various types of multimedia data and to model complex information structures.

This paper suggests an indexing technique of video materials in the compressed format itself in order to reduce the cost and load which will be caused at the time of indexing the compressed video information.

With rapid growing computer and communication technologies the type of digital media we face have changed from text to multimedia such as image and video. In this research, we had study contents-based multimedia information retrieval technologies

목 차

I. 서 론	1
II. 기반 기술	3
2.1 멀티미디어 시스템 요소 기술	3
2.2 부호화 기술의 필요성 및 원리	3
2.3 부호화의 성능 평가	4
2.4 부호화 기법 및 요소 기술	6
III. 내용기반 정보 검색	16
3.1 비디오 정보와 응용	16
3.2 비디오 처리 기술	17
3.3 비디오 색인	18
IV. 내용기반 정보 알고리즘	23
4.1 내용기반 검색 알고리즘	23
4.2 MB 기반 컷 검출 알고리즘	28
4.3 특징들을 이용한 비디오 인덱싱	31
V. 실험 및 결과 고찰	33
VI. 결 론	36
參 考 文 獻	37

I. 서 론

디지털 압축기술의^[3] 발달과 컴퓨팅 능력이 발전함에 따라서 하루에도 많은 양의 이미지, 그래픽, 오디오, 비디오 정보가 생성되고, 사용되고 있다. 더욱이 인터넷의 폭발적인 성장으로 이러한 멀티미디어 정보는 더 이상 특정 지역에 국한되지 않는다. 가까운 장래에 네트워크를 통하여 다양한 멀티미디어 콘텐츠가 유통될 것이며, 전세계에 분산되어 있는 멀티미디어 콘텐츠에 컴퓨터 혹은 멀티미디어 단말기를 통하여 액세스할 수 있을 것이다 그러나 아직까지는 멀티미디어 정보를 문자처럼 빠르고 정확하게 찾기가 어렵다. 멀티미디어 콘텐츠를 단순한 키워드나 파일 이름으로 색인하여 정보를 검색하는 방식은 폭발적으로 증가하고 있는 다양한 멀티미디어 정보를 검색하는 데에는 한계가 있기 때문이다. 따라서, 멀티미디어 정보를 문자 정보처럼 정확하고, 신속하게 찾는 것은 다양한 멀티미디어 응용 분야에서 요구되고 있는 중요한 기술로서 많은 요구가 활발히 진행되고 있다.

또한 멀티미디어의 중요성을 볼 때, 의미를 갖는 이유는 저장 장치, 통신, 데이터 압축 기법 등의 발달로 다양한 형태의 멀티미디어 응용분야가 가능해지고 있다는 점이다. 여러 종류의 멀티미디어 응용 중 주문형 비디오(Video On Demand, VOD)서비스는 기존의 TV 유선 방송을 대체할 수 있는 획기적인 서비스 분야로 각광을 받고 있다.

본 논문에서는 멀티미디어의 핵심 기술에 있어서 비디오 데이터의 검색 요청에 대해 데이터를 검색하여 사용자에게 공급해주는 디지털 비디오 데이터 검색에 대해 기술한다.

또한, 데이터 검색을 위하여 키워드를 선정하듯이 디지털 비디오 검색에도 키 프레임이 선정되거나 물체가 추출되어 이를 바탕으로 인덱싱이 이루어져야 한다.^[5] 그러나 디지털 비디오는 대용량이기 때문에 압축을 피할 수 없고, 압축 상태에서는 장면의 특징, 등장 요소 등이 전혀 보존되지 않아서 인덱싱을 위한 키를 선정할 수 없다. 그러므로 압축 비디오는 원래의 상태로 환원되어 인덱싱된 후 재차 압축되어야 한다.

본 논문에서는 대용량의 비디오 데이터에서 정확하고 효율적인 검색을 위해 MPEG-2 Video 상에서 특정 물체를 추출해 낼 수 있는 방법인 색상을 이용한 물체의 추출에 대해 연구한다. 이러한 색상을 이용한 물체의 추출 방법은 히스토그램 기반의 장면 변화를 검출하기 위해 색상 분포에 의한 칼라 영역의 자동 추출, 효율적인 인덱싱 그리고 효과적인 검색이 동시 만족을 위해 사용되어 진다.^{[5][8]}

본 논문은 멀티미디어 저작 및 비디오 생산은 비디오 생산 절차를 설계하기 위한 응용 예로 사용될 수 있으며, 멀티미디어 세분화는 특징에 기반 한 분류로서 비디오 삽입의 첫 번째

과정이며 비디오 인덱싱은 두 번째 과정이며, MPEG-2 Video상에서의 내용 기반 검색에 대해 연구하며, Video 데이터 상에서의 비디오 데이터 모델링 및 인덱싱 기술에 관해 논하였다.

본 논문의 구성을 살펴보면, 2장에서는 멀티미디어 데이터와 시스템에 있어서 사용되어지는 요소와 멀티미디어 데이터로 구성되어진 부호화 기술과 보다 더 효율적인 데이터의 구성을 위한 압축 기술에^[7] 대해 기술하였고, 3장에서는 내용기반 정보 검색을 위한 비디오 데이터의 처리 기술과 색인을 위한 정보에 대해 기술하며, 디지털 방식으로 저장된 비디오 데이터 상에서 내용기반 검색을 위해 사용되는 검색 알고리즘인 색상 히스토그램 기법을 이용하여 영상 검색을 위한 색상 컬러 특징 추출 방법과 영상의 컬러, 밝기, 에지 성분을 가지고 영상의 중요한 특징을 효과적으로 표현할 수 있는 Composite(Color/Edge) 특징 추출 방법을 제안하며, 주요 컬러의 매칭 효과를 제공할 수 있는 RGB의 3차원 공간상의 점들의 정규화를 위한 방법을 제안한다. 또한 4장에서는 MPEG-2 Video 상에서 압축된 영역의 특징물을 추출 및 표현에서 인덱싱과 검색 단계를 실험을 통해 입증한다.^[8]



II. 기반 기술

2.1 멀티미디어 시스템 요소 기술

메모리 IC의 급격한 발전과 반도체 기술인 메모리와 마이크로프로세서를 중심으로한 디지털 신호 처리용 반도체(Digital Signal Processor: DSP)의 기술, 부호화 기술과 통신 기술의 발전, 여러 개의 시스템이 통신망에 연결되어 정보의 수신, 수신 공유가 가능한 네트워크 기술의 발전 등으로 멀티미디어 시스템 기반 조성이 되고 있다. Fig.1은 이러한 기반기술과 멀티미디어 시스템의 관계를 나타내고 있다.

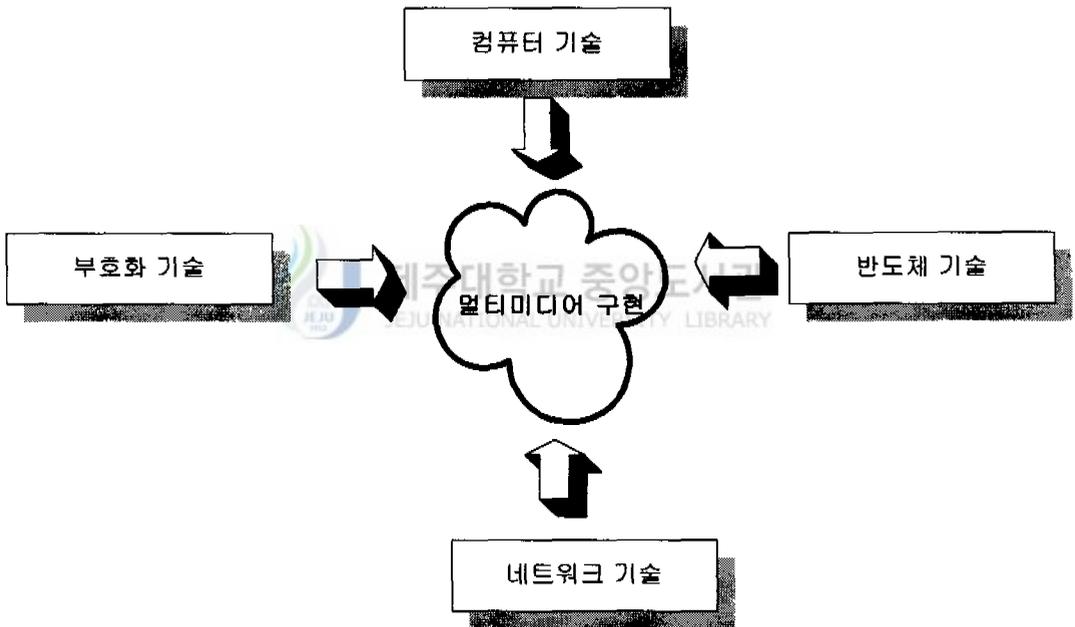


Fig.1 멀티미디어 시스템 요소 기술

2.2 부호화 기술의 필요성 및 원리

숫자나 문자로 된 텍스트 데이터들로부터 정보를 얻어내는 기존의 방식 대신, 오디오나 이미지 등 좀더 인간의 감각에 접근한 형태로 표현된 데이터들로부터 직관적으로 정보를 얻어 내려는 욕구가 강해졌다. 이러한 요구는 관련된 기술들, 즉 센싱, 신호처리, 통신, 디스플레이

이에 관련된 기술과 이들의 기초가 되는 소재 및 반도체, 그리고 소프트웨어 기술의 결합에 힘입어 현실 응용화 되는 등 날로 총족되어 가고 있다. 따라서 이들을 저장, 처리, 전송하기 위해서는 이들은 부호화하는 기술이 필연적이다 이러한 부호화 기술에는 다음과 같은 기본 원리가 적용된다.^{11)[2]}

가. 중복성 제거의 원리

불필요한 정보를 삭제하고 반복되는 정보를 간략히 기술함으로써 데이터의 크기를 줄이는 원리이다. 전화 음성의 경우 말을 하는 때와 말을 하지 않은 때가 분리되는 점을 이용해서 말을 하지 않을 때는 정보를 보내지 않음과 동영상의 경우 1초에 비슷한 그림을 60장 정도 빠르게 바꾸어줌으로써 움직이는 그림을 구현하기 때문에 직전 그림과 현재 그림이 별 차이가 없다는 점을 이용하여 정보의 양을 줄인다.

나. 가시성·가청성의 원리

가시성·가청성의 원리는 사람이 감지할 정보만을 추려서 남기고, 감지 할 수 없는 정보를 삭제함으로써 데이터의 양을 줄인다.

다. 명도 정보는 자세히 묘사하고 색상은 덜 자세히 묘사

사람이 그림을 볼 때 명도의 변화에는 민감한 반면 색상의 변화에는 둔감한 특성이 있다. 이를 이용하여 명도정보는 자세히 하되 색상에 대해서는 덜 자세히 묘사함으로써 정보의 양을 줄일 수 있다.

2.3 부호화의 성능 평가

부호화 시스템을 비교하기 위해서는 이를 평가할 수 있는 방법이 있어야 한다. 이를 위해서 부호화 정도, 영상의 품질, 부호화/복호화 속도 그리고 부호화/복호화를 위하여 특별한 하드웨어, 소프트웨어가 있는가를 고려해야 한다.

가. 부호화율

부호화의 성능은 부호화 과정의 입력과 출력된 데이터 양의 비율로 나타낸다. 그러나 이 방법은 입력된 것과 출력된 것의 데이터 형식이 다를 경우 정확한 척도라고 볼 수 없다. 예를

들어 512×240 해상도, 24bpp(bits-per-pixel) 영상을 입력으로 받아 256×240 해상도의 영상을 15,000 바이트로 출력한 부호화 시스템이 있다. 부호화율은 단순히 입력/출력 데이터의 크기 비율로 나타낼 경우, 이 예와 같은 부호화 시스템은 입력이 512×480×3 = 737,280 바이트이므로 737,280:15,000 = 49:1의 부호화율을 제공한다. 그런데 부호화 시스템은 영상 해상도의 감소, 즉 영상의 크기를 작게 함으로써 영상을 표현하기 위하여 사용된 저장공간의 크기를 (512×480):(256×240) = 4:1로 감소시켰다. 입/출력 영상의 해상도가 같을 경우 이 부호화 시스템은 대략 12:1의 부호화율을 제공할 것이다. 이 예에서와 같이 단순히 입/출력 데이터 양의 비율은 출력된 영상의 해상도가 같지 않을 경우 부호화 성능의 정확한 척도가 될 수 없다 앞서 언급한 방법보다 더 좋은 것은 픽셀당 소요되는 데이터의 양을 비교하는 것이다. 예를 들어 256×240 해상도와 24bpp의 영상을 상영하기 위하여 15,000바이트가 사용되었다면 출력된 영상은 (15,000×8)/(256×240) = 2bpp를 사용하므로 12:1의 부호화율을 나타낸다.

나. 영상의 품질

영상의 품질에 대하여 논할 때 부호화를 무손실 부호화와 손실 부호화로 나누어 생각한다. 원천 영상의 표현을 효율적으로 하여 사용하는 저장 공간을 감소 시킬 뿐 영상의 정보를 손실하지 않는다. 무손실 부호화된 영상은 원천 영상과 정확하게 일치한다. 반면에 손실 부호화는 원천 영상과 정확하게 일치하지는 않으며 복원시 사람이 인식하지 못할 정도의 근소한 차이만을 가진 영상을 상영할 수 있도록 하는 것을 목적으로 하고 있다.

다. 부호화/복호화 속도

많은 응용에서 부호화와 복호화는 다른 시점에서 발생한다. 심지어 다른 장소, 완전히 다른 시스템에서 발생할 수 있다. 이러한 이유로 부호화와 복호화 과정에서 영상의 저장과 전송이 있으며, 우선적으로 부호화를 필요로 하고 있다 따라서 부호화와 복호화의 속도는 별도로 평가할 필요가 있다.

라. 특별한 하드웨어/소프트웨어의 필요 유무

어느 정도의 부호화와 복호화는 컴퓨터에서 소프트웨어로 가능하다. 그러나 단순한 알고리즘은 부호화율이 좋지 않으며, 복잡한 알고리즘은 입력된 각 픽셀당 수십, 수백 심지어 수천 개의 연산을 필요로 하기 때문에 소프트웨어적인 압축은 많은 시간을 필요로 한다. 이것은 컴퓨터의 발전에 따라 해결될 수도 있지만 현재 대부분의 시스템은 특별한 하드웨어의 도움으로 속도를 향상시키고 있다.

2.4 부호화 기법 및 요소 기술

부호화는 여러 가지 응용에 따라 다양한 기법이 존재하지만 크게 무손실 부호화 기법과 손실 부호화 기법으로 분류할 수 있다. 본 절에서는 이와 같은 부호화 기법을 영상 데이터 부호화를 중심으로 기술하였다.

가 무손실 부호화 기법

무손실 부호화 기법이란 부호화한 데이터를 복원했을 때 복원한 데이터가 부호화 전의 데이터와 완전히 일치하는 것을 말한다. 이 기법은 부호화 한 데이터가 부호화 할 데이터에 어떤 변경이나 수정은 가하지 않는다. 따라서 멀티미디어 정보에서 정확성을 요하는 의료 영상의 부호화에 사용할 수 있다. 무손실 부호화의 경우 복원시 원래의 데이터와 동일한 정보를 얻을 수 있지만, 손실 부호화 기법에 비해 부호화율이 낮다는 단점이 있다.

(1) 허프만 부호화(Huffman Coding)

허프만 부호화는 일반 응용에서 가장 많이 사용되며, 기본 개념은 발생확률이 높은 심볼에 대하여는 짧은 길이의 부호어를 발생 확률이 낮은 심볼에 대하여는 상대적으로 긴 부호어를 할당하여 평균 부호어 길이를 짧게 하는 것이다

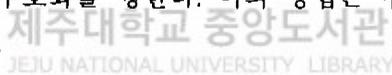
Fig.2 는 심볼이 5인 허프만 부호화의 작성의 예를 나타내고 있다.

심볼	확률	순서	부호어
0	0.42		1
1	0.20		000
2	0.15		001
3	0.15		010
4	0.10		011

Fig.2 허프만 부호화의 예

우선 심볼의 출현 확률이 작은 것을 2개 선택하여 2개중 발생 확률이 작은 것에 1을 큰 것

에 0을 할당하고 나서 2개를 통합하여 하나의 심볼로 만든다. 이 때 새로 만들어진 심볼의 발생 확률은 2심볼의 합이 된다. 이와 같이 하여 처음 5개의 심볼은 4개로 된다. 다음에 4개의 심볼을 다시 출현 확률순에 따라 2개를 선택한 후, 발생 확률이 작은 것에 1을 큰 것에 0을 할당하고 2심볼을 통합하여 하나의 심볼로 만든다. 이와 같은 처리를 심볼수가 2개 남을 때까지 반복한다. 할당된 부호를 역순으로 호출하여 그것에 대응시키면 그것이 심볼의 부호어가 된다. 이러한 허프만 부호는 부호화 방법이 간단하고 소스 엔트로피에 비교적 가까운 평균 비트율을 가지나 몇 가지 기본적인 제약을 갖고 있다. 첫 번째 제약은 이산 무기역 소스의 임의의 심볼 S_i 에 대한 이상적인 이산 부호어의 길이는 $-\log_2 p^{(s_m)}$ 이다. 부호어는 반드시 정수 값으로 표현되기 때문에 이는 소스 심볼의 확률이 1/2, 1/4, 1/8 등과 같이 2의 음의 멱급수(Negative Power of Two)를 가져야 함을 의미한다. 따라서 앞에서 설명한 바와 같이 이산 무기역 소스의 임의의 심볼이 이상적 이산 부호어 길이와 같아지기 위해서는 심볼의 발생확률이 2의 음의 멱급수로 표현되어야 하나 실제 응용에서는 불가능하다. 또 다른 제약으로는 소스의 발생 확률이 변화되었을 때 허프만 부호는 효율적으로 대처하기 어렵다는 점이다. 실제로 사전에 고정된 허프만 부호를 임의의 다른 소스에 적용한 경우 오히려 데이터가 증가하는 경우도 있다. 이와 같은 결점을 해결하기 위해서 허프만 부호는 종종 2패스(Path) 기법을 사용하기도 한다. 즉 먼저 심볼의 통계적 성질을 구하여 부호어를 발생시킨 후, 이를 근거로 실제 부호화를 행한다. 이의 방법은 부호화율은 상승하나 부호화를 2번 행해야 하는 결점이 있다.



(2) 산술 부호화(Arithmetic Coding)

허프만 부호의 부호화 효율을 높이기 위한 방법의 한 가지로 소스 열을 몇 개의 블록으로 나눈 후, 각 블록을 가변장 부호어로 할당하는 것이다. 물론 복호기 측에서는 각각의 부호어에 대응하도록 가변 블록으로 분리한 후, 원래의 부호로 변환한다. 즉 소스열 블록과 부호어 블록사이에는 일대일 대응 관계가 존재한다. 이에 비해 산술부호는 비블릭 부호(또는 트리(Tree)부호)로 길이 m개의 심볼로 구성되는 전체 입력 데이터 열에 단 하나의 부호어를 할당하는 기법이다. 산술부호에서 부호어 길이는 대략 이상적 이산 부호어 길이 $-\log_2 p^{(s_m)}$ 에 근접하며, 여기서 $p^{(s_m)}$ 은 소스열(s_m)의 발생확률을 나타낸다. 산술부호의 기본개념을 이해하기 위하여 이산 무기역 소스로 발생되어진 2진 심볼열(s_m)의 부호화를 고려해 보면 우선 심볼열에서 '0'이 나올 확률을 p라하면 '1'이 나올 확률은 q=1-p가 된다. 여기서 왼쪽값은 포함하고 오른쪽 값은 포함하지 않는 구간(0,1)을 I라 정의하기로 한다. 한편 길이 m의 소스열(s_m)에 있어서 발생할 모든 조합은 2^m 이 되며, 이의 발생 확률 $p^{(s_m)}$ 의 총 합은 1이 된

다. 따라서 서브(Sub) 구간, $I_l, l=1,2,3,\dots, 2^m$ 은 각 소스열 s_m 에 일시적으로 할당되어 질 수 있으며, I_l 의 길이는 $p^{s(m)}$ 과 같으며 서브구간은 중첩되지 않는 방법이다

(3) 반복길이 부호화(Run-Length(RL) Coding) 기법

이 기법은 반복되는 픽셀들을 하나의 색상 값과 그 색상이 반복되는 횟수로 치환하는 것으로 동일한 값을 가지는 픽셀들이 인접하게 모여있는 특성을 부호화 방법에 이용한 것으로써 영상의 종류에 따라 부호화율이 크게 달라진다. RL 부호화 방법은 총 16횟수의 영상의 표현은 색상 값과 그 색상이 연속적으로 나타나는 횟수의 쌍으로 표현되고 있다. 횟수는 0번이라는 것이 없기 때문에 '0000'이 1번 발생을 의미하고 '1111'은 16번의 반복을 의미한다.

나. 손실 부호화 기법

손실 부호화 기법은 높은 효율의 부호화율을 얻기 위하여 부호화시 화질 열화를 허용하는 기법으로 기본 구성은 Fig.3 과 같다.

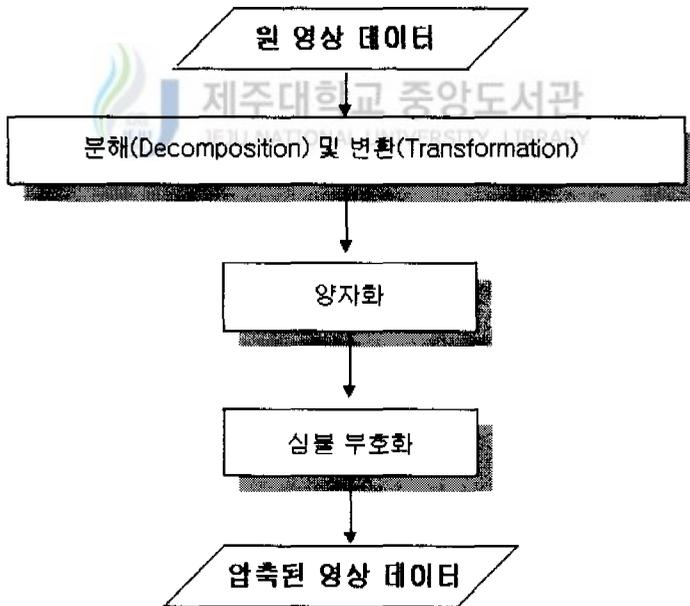


Fig. 3 손실 부호화의 일반적인 구조

이와 같은 손실 부호화 기법으로는 색상표(Color Lookup table)를 이용하는 방식, 보정(Interpolative)기법, 예측 부호화(Predictive Coding)기법, 변환 부호화(Transform Coding)

기법 등이 있다

(1) 색상표를 이용하는 방식

각 픽셀의 값은 실제 표현될 색상을 나타내는 것이 아니라 색상표의 색인이다. 즉 픽셀의 색상은 픽셀의 값과 색상표에 의해서 결정된다. 이 방법은 색상표의 크기에 따라서 부호화율이 결정된다. 예를 들어 24bpp로 된 원천 영상을 256개의 색상을 가진 (즉 8bpp) 색상표를 이용하여 변환한다면 부호화율은 3:1이 된다 이 방법은 영상 전체에서 사용할 수 있는 색상을 256개 혹은 더 적은 수의 색상으로 표현하는데 주로 이용된다.

(2) 보정(Interpolative) 기법

보정 부호화 기법은 영상을 구성하는 픽셀들의 부분 집합만을 전송하고 픽셀들 사이에 생략된 픽셀은 보정에 의해서 생성하는 기법을 말한다. 이 기법은 생략된 픽셀의 값을 그 주변 픽셀의 값을 이용하여 추정하여도 사람이 인식하기 어려울 때 사용한다.

(3) 예측 부호화(Predictive Coding) 기법

일반적인 화상의 경우 인접 픽셀 값 사이에는 큰 상관성이 존재한다. 따라서 임의의 한 픽셀 값에 대한 많은 정보를 인접 픽셀 값으로부터 얻는 것이 가능하다. 예측 부호화 기법에서는 위와 같은 성질을 이용하여 부호화를 행하는데, 부호화의 기본 개념은 원하는 픽셀값의 예측은 주위의 픽셀 값을 이용하여 행하는 것이다. 예측을 기반으로 한 부호화 기법들로 는 차이 PCM(Differential PCM: DPCM) 과 적응적 DPCM(Adaptive PCM. ADPCM)등이 있다

(4) 변환 부호화(Transform Coding) 기법

일반적인 변환 부호화 기법은 먼저 $N \times N$ 크기의 원 영상을 보다 작은 $n \times n$ 크기의 서브블록으로 분할한 후, 각각의 서브 영상에 대하여 유니터리(Unitary) 변환을 행한다. 변환의 목적은 원 신호 사이의 상관성을 제거하는데 있으며, 변환 결과는 변환 계수중 일부 계수에 에너지가 집중되는 현상을 보인다. 이와 같은 방법에 의해 에너지가 거의 없는 변환 계수는 양자화시 제거되며, 결과적으로 부호화가 가능하다. 또한 사람 눈의 특성을 이용하여 눈으로 보기에는 왜곡이 없는 압축을 얻는 것도 가능하다. 변환 부호의 기본 구성은 Fig.4 에 도시하였다.

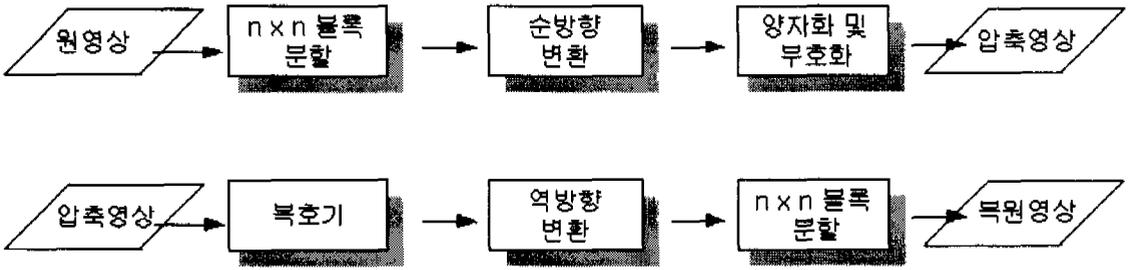


Fig.4 변환 부호의 기본 구성

영상 부호화를 목적으로 사용되는 변환은 다음과 같은 몇 가지 특성을 가져야 한다.

- 1 영상 데이터의 상관성을 제거하는 것
- 2 영상과 무관한 기저함수(Basis Function)를 가질 것
- 3 구현이 용이할 것(처리속도가 빠를 것)

변환 부호에는 좌표축 회전에 의한 변환이나 기본 함수 분해에 의한 변환이 있으나 여기서는 주로 많이 사용되는 영상 변환에 대해서만 설명하도록 한다.

(가) MPEG-2



MPEG-2 비디오는^[2] 일종의 범용 압축 알고리즘으로, MPEG-1 비디오를 크게 확장 발전 시키면서 많은 도구들을 마련하여 응용분야에 따라 이들을 적절히 선택 사용하도록 하고 있다. 압축효율의 향상을 위해 MPEG-2 비디오 필드단위의 처리, 움직임 추정과 보상방식, 양자화, DCT 계수와 주사방식 그리고 가변장 부호화 등 많은 부분들이 개선되었다.

MPEG-2 비디오의 범위는 매우 넓지만 응용분야마다 그 중 특정 해상도에 특정 기능까지만을 사용한다. 따라서 부호기(Decoder)와 복호기(Encoder)를 제작할 때의 편의를 위해 MPEG-2 비디오를 해상도와 기능에 따라 몇 가지로 분류하고 있다. 우선 화면의 해상도는 4개의 레벨로 분류된다.

많은 표준화 작업을 통해 MPEG은 DCT 부호화와 예측부호화를 이용하여 상당한 압축된 정보를 얻을 수 있다 그 데이터 구조를 살펴보면 Fig 5와 같이 시퀀스 층, GOP층, 픽처 층, 슬라이스 층, 매크로블록 층, DCT 계수를 포함하는 블록 층으로 분류되어진다.^[2]

1) 시퀀스 층

비디오의 수평크기와 수직크기, 비디오내의 매크로블록 수, 초당 매크로블록 수, 화면을,

순방향과 역방향 f코드, 버퍼크기: VBV(Video Buffering Verifier), 비트율 등이 지정된다.

2) GOP 층

몇 개의 프레임들을 일련의 집합으로 표현한 것으로 비디오 정보의 랜덤액세스에 이용된다. GOP내의 프레임 수(N)나 I 또는 P 프레임이 나타나는 주기(M)를 지정한다. 비디오 스트림상에서 GOP의 최초는 I 프레임이고, GOP의 마지막은 I 또는 P 프레임이 된다. 통상적으로 이용되는 프레임 수(N)는 15이고 주기(M)는 3이다, 즉 I와 P 프레임 사이에 두 개의 B 프레임이 존재한다.

3) 픽처 층

한 프레임에 대한 픽처타입, 움직임벡터에 대한 플래그, 확장비트 등의 공통적인 속성을 지정하는 층이다.

4) 슬라이스 층

한 프레임을 임의의 길이(MB들의 집합)로 분할한 것을 슬라이스라 하며, 이에 대한 공통적인 정보를 지정하는 층이다. 슬라이스는 임의의 길이의 매크로블록(MB)의 띠이며 여러 픽처에 걸쳐 있을 수는 없다. 처음과 마지막 MB는 서로 연계할 수 없고, 한 개의 MB로 구성된 슬라이스의 경우 그 MB는 건너 뛴 수 없다.

5) 매크로블록 층

슬라이스 층을 더욱 분할하여 매크로블록에 대한 공통적인 정보를 지정한다. 부호 발생량이 적을 때 삽입하는 매크로블록 스테핑, 매크로블록 Escape, MBA(Macroblock Address), MB 타입 등이 지정된다.

6) 블록 층

매크로블록 층을 8×8블록으로 더욱 세분화한 층이며, 해당되는 DCT 계수를 포함하고 EOB(End Of Block)로 종료된다. 인트라(Intra) DC는 독자적인 VLC(Variable Length Coding : 가변장 부호화)를 사용한다.

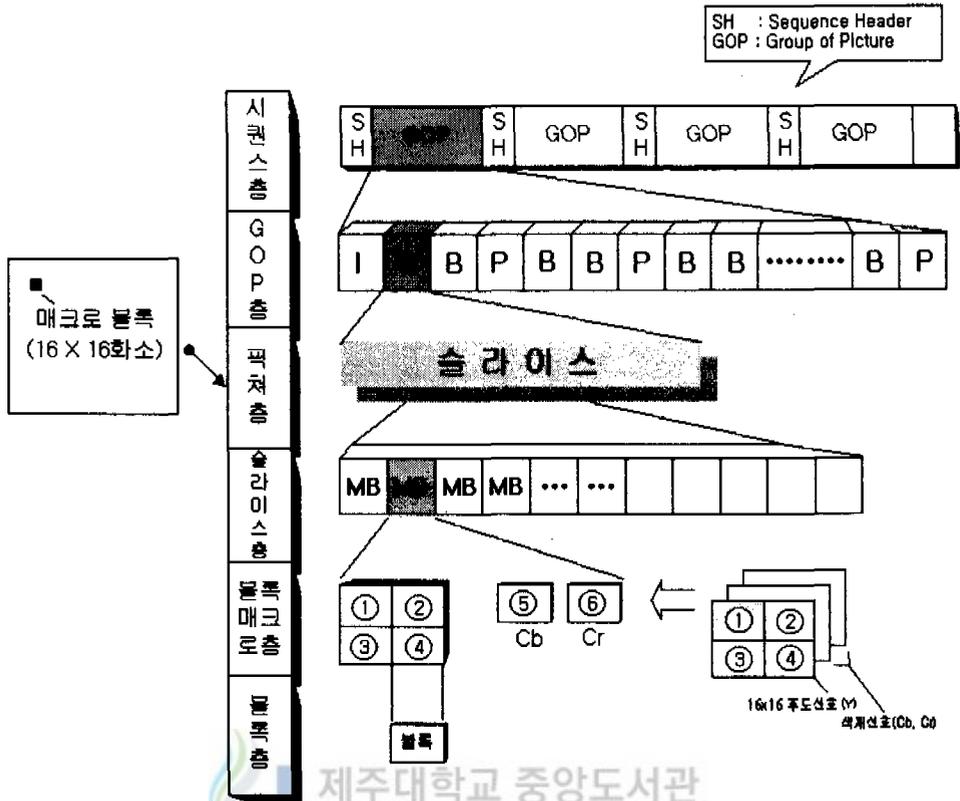


Fig. 5. MPEG-2 Video의 구조

(나) DCT

DCT는 인접 화소간의 값이 유사한 영상의 상관관계를 이용하여 정보를 부호화하는 기법으로 Fig.6 에서 처럼 원 영상을 8x8 픽셀 블록들로 나누어 각 블록에 대해서 푸리에 변환을 통해서 영상을 공간 영역에서 주파수 영역으로 변환한다.

변환 전에 불규칙적으로 퍼져있던 화소 값이 변환 후에는 저주파 향 쪽으로 큰 값이 집중되며 상대적으로 고주파 향 쪽으로는 작은 값이 집중된다. 대부분의 데이터가 낮은 주파수 향 쪽으로 집중됨으로 데이터가 거의 없는 고주파 향 쪽은 무시하고 부호화를 행하여도 높은 부호화 율로 우수한 화질을 얻을 수 있다.

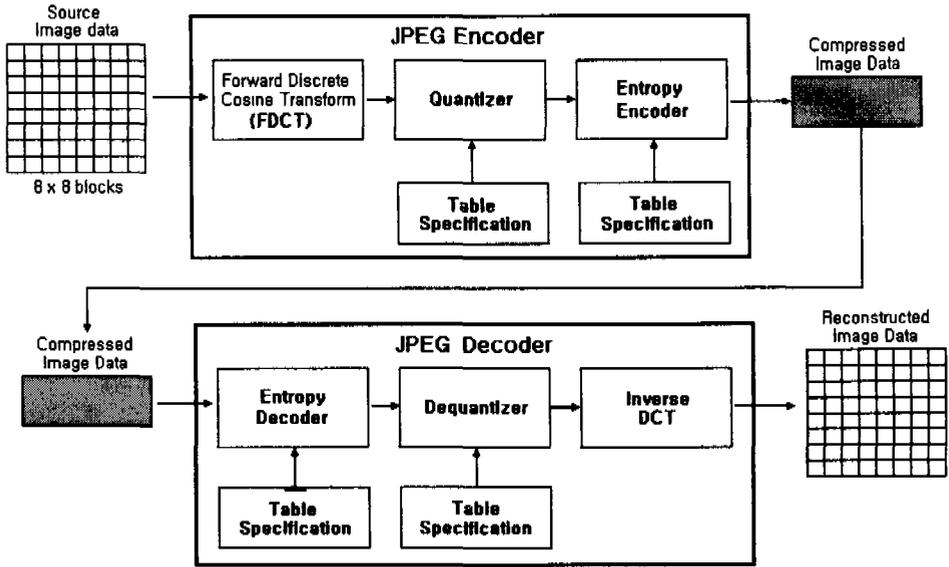


Fig.6 DCT 과정

(다) Wavelet 변환

웨이블릿 변환은 신호를 다해상도(Multiresolution)로 분해하며, 입력 신호는 서로 다른 해상도로 표현하기 위한 부 신호들과 나머지 신호들을 표현하기 위한 세부 신호들로 나누어진 다. 영상이나 음성 등의 신호들은 에너지 분포가 주파수 영역에서 관찰할 때 고르지 못한 분포를 가지고 있다. 이로 인해 신호를 분해하는 것이 필요하게 되었으며, 기본적으로 도입된 개념이 스펙트럼을 부 밴드로 분할하여 각각의 부대역을 독립적으로 목적에 맞도록 처리하는 M대역 분해이다. 신호를 다해상도로 분해하는 것은 두 가지 방법이 있다. 하나는 M개의 동일한 대역(Equal Bandwidth)으로 분할하는 방법이고, 다른 하나는 Dyadic 트리 구조로의 분해인데 웨이블릿 변환이 이에 속한다. 웨이블릿으로 변환된 영상 신호는 서로 다른 주파수 특징을 갖고 이들간에는 어느 정도의 상관관계가 존재한다. 웨이블릿 변환을 이용하여 영상을 다해상도로 분해한 후 영상을 부호화하는 것은 기존의 DCT를 이용한 방법보다 부호화율을 높일 수 있고 블록킹 현상이 일어나지 않아 인간의 시각에 보다 적절한 부호화 방법이다.

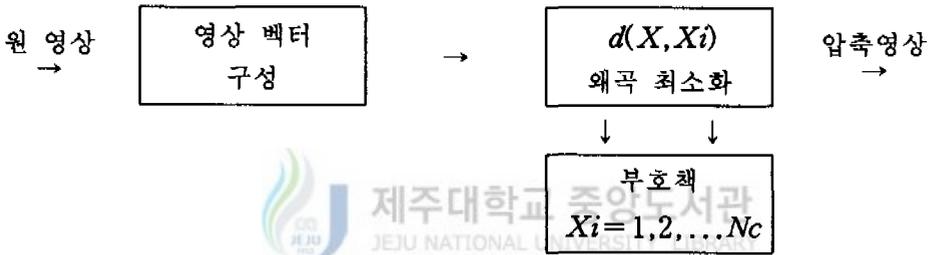
인간의 시각은 세밀한 부분보다 불연속성이나 갑작스런 밝기의 변화와 같이 해상도가 낮은 부분의 영상에 대해 더 민감한 반응을 보인다. 웨이블릿 변환으로 분해된 부대역 영상들은

이러한 인간의 시각적 특성을 잘 반영하고 있어 영상을 효율적으로 부호화 할 수 있다. 즉 영상신호를 웨이블릿 변환을 이용하여 다해상도로 분해하고 인간시각 특성을 고려하여 추출된 에지정보를 이용하여 분해된 부밴드 영상을 양자화 한 후 양자화된 계수들의 통계적 특성과 상관관계를 이용하면 효율적으로 부호화 할 수 있다.

(5) 벡터 양자화(Vector Quantization) 기법

벡터 양자화 기법은 일련의 데이터로 구성되는 벡터를 양자화하여 부호화하는 기법이다. 벡터 양자화는 스칼라 양자화보다 부호화율이 더 높으며, 복호기도 비교적 간단하다. 기본적인 벡터 양자화기의 블록도를 Fig.6 에 나타내었다.

부호기



복호기

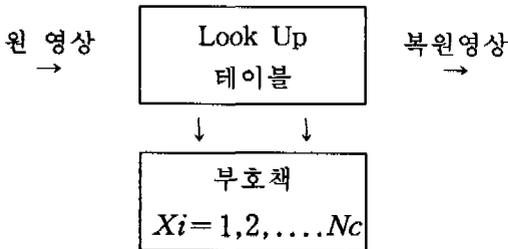


Fig.7 벡터 양자화기 블록도

원 영상은 먼저 n-차원의 영상 벡터로 분해되어지는데 영상 벡터를 구하는 방법은 응용에 따라 서로 다르게 구해진다. 다음에 이와 같이 얻어진 영상 벡터 X와 부호책(Codebook)내에 있는 부호 벡터 $X_i, i=1, \dots, N_c$ 를순차적으로 비교하여 왜곡이 최소화되는 부호 벡

터를 찾는다. 최소 왜곡 부호벡터가 선택되면, 그것의 인덱스 k 가 $\log_2^{N_r}$ 비트를 사용하여 전송된다. 복호기에서는 부호기와 동일한 부호책으로 부터 지표 k 에 해당되는 부호벡터를 찾아서 원영상으로 복원한다.

벡터 양자화에서의 효율은 부호책에 크게 의존하는데 Linde-Buzo-Gray(LBG) 구성방식이 가장 많이 알려져 있다. 벡터 양자화에는 부호책의 크기선정이 중요한데 효율을 높이기 위해 대표 벡터의 수를 늘리면 부호책이 커져 부호기에서 대표벡터를 선정하기 위한 탐색시간이 길어져 실시간 수행이 어렵다. 따라서 계산 복잡도 및 메모리에 요구사항을 줄일 수 있는 다양한 방식에 대한 연구가 많이 진행되고 있으며 다른 부호화 기법과 결합한 벡터 양자화가 많이 연구되고 있다.



Ⅲ. 내용기반 정보 검색

멀티미디어 기술은 디지털 방식으로 비디오 데이터의 저장 및 처리를 가능하게 하고 있다. 또한, AVID Media Composer와 같은 디지털 비디오 편집 시스템은 사용자에게 디지털 비디오 부분들을 조합해 새로운 디지털 비디오 항목으로 만들 수 있게 하여 준다. 디지털 비디오는 정보의 질과 비디오 데이터의 접근성을 좋게 할 수 있다. 그러나 이러한 디지털 비디오 기술만 가지고는 모든 문제를 해결할 수 없다. 즉, 비디오 정보의 일관성있는 공유, 재사용성 및 검색과 관련하여 비디오와 메타 데이터를 통합 제어 할 수 있는 시스템이 필요하며 이는 전통적인 데이터베이스 시스템의 필요성과 유사하다. 비디오 데이터관리 시스템은 데이터 및 메타 데이터를 안전하고 일관성 있게 최소한의 중복으로 저장하여야 하며 사용자가 비디오 데이터형식에 구애받지 않는 데이터 독립성을 제공하여야 한다. 이러한 비디오 관리 시스템은 비디오 사용성이 내용 기반 접근이 가능하도록 하는 새로운 차원으로 발전됨에 따라 내용 기반 접근기능을 제공할 수 있도록 비디오 데이터베이스를 관리하여야 한다. 내용 기반을 허용하는 비디오 데이터 관리는 데이터 모델링, 비디오 삽입과 구성 및 검색 과정이 기존의 데이터베이스 관리 시스템과는 전혀 다르다.^[6]

비디오 데이터 모델은 전형적인 데이터 접근 형태를 쉽게 하기 위해 설계된 응용 분야의 특성에 맞는 표현 방법이다. 따라서 데이터 모델을 설계하기 위해서는 비디오 응용 분야들의 특성을 분석하여야 하며 분할(Segment)된 비디오에 대한 메타 데이터를 이용한다.

3.1 비디오 정보와 응용

전통적인 영화, 아날로그 비디오 그리고 디지털 비디오는 다른 기술들을 의미하지만, 이러한 내용은 동영상(Motion Picture) 자료형과 같은 데이터형의 변형이다. 비디오 정보는 매우 설득력 있게 정의된 용어는 아니나, 물리적 단계에서 비디오는 이미지의 스트림이다. 비디오는 여러 샷(shot)들과 오디오 레코딩들로 구성된 복잡한 멀티미디어 문서이다.^[4] 비디오 문서(video document)는 비디오 구성을 의미하는 용어로서 사용된다. 비디오 문서는 정지한 이미지, 오디오 그리고 비디오 데이터와 같은 미디어 데이터로 정의된 합성물이다. 비디오 정보는 메타 데이터 그리고 비디오 데이터와 연관된 메타 데이터를 포함하는 집합적인 용어로

사용된다.

3.2 비디오 처리 기술

비디오(미디어)처리 기술 특히, 디지털 비디오 처리 기술은 종래의 수동적인 시청 방식에서 탈피하여, 새로운 패턴을 제시한다. 즉, 사용자가 단순 조작에 의해 비디오를 시청하는 방식이 아닌 원하는 비디오의 내용을 원하는 형태의 모습으로 시청할 수 있게 한다. 이와 같이 사용자가 원하는 형태로 비디오를 다루기 위해서 비디오의 처리 방식에 있어서도 새로운 기술들이 제시되고 있다.

가. 비디오 처리 방식(구조적인 측면)

① 비디오 스트림 분석

② 자동으로 비디오 스트림을 샷으로 구분

③ 각 샷은 카메라 동작에 관한 정보와 표현되는 색상과 같은 추출된 특징들을 사용하여 색인하며, 이러한 과정을 비디오 구조화라 하고, 그 결과로 기술된 내용을 구조화된 비디오라 한다. 여기서의 카메라의 동작에 관한 정보의 표현 및 검색은 스테레오 영상에 의한 카메라 교정(Calibration)을 통한 방법과 정보의 표현과 색상 정보, 색상 변환에 의한 방법은 RGB 모델의 3차원 공간상의 점들의 정규화 과정에 의한 검색 방법에 의한 연구를 서술한다.

나. 비디오 구조

비디오를 샷으로 나누는 방법으로는 컷(Cut)를 판별하는 방법이 있다. 컷 판별은 새로운 연구 분야는 아니지만, 비디오 컴퓨팅에 있어서 컷 판별은 가능한 정확해야 한다. 컷 판별 알고리즘에서는 비디오 프레임간의 색상, 밝기, 움직임 방향 등의 차이(인터프레임 차이)로서 비연속성을 감지하여 컷을 찾아낸다. 비디오에서의 물체의 급격한 이동이나 Telecine conversion, 느린 동작, 애니메이션, 플래쉬 조명(Strobe lighting)등은 인터프레임 차이가 매우 크기 때문에 탐지되지 않는 경우도 있다.

(1) 샷 분석

- 일반적으로 샷은 카메라의 조건과 장면에 있는 대상 물체에 대한 정보를 가진다.
- 장면 조건이 특정 용용에 국한되지 않고서는 컴퓨터에서 각 객체를 인식하기란 쉽지 않다.
- 샷은 시공간적으로 투시된 이미지를 포착한 표본화된 프레임의 연속이므로, 장면의 시, 공간적인 개념을 가지고 있다.
- 따라서 다음과 같은 두 가지의 방법으로 이러한 정보를 추출한다.
 - ① camera work information : Scene의 공간적인 상황 파악
 - ② Representation color information : 대상에 대해 필요한 정보 파악
- 카메라의 이동에 따라서 포착된 물체는 전체적인 변화를 가져온다.
- 카메라를 오른쪽으로 움직이면(Pan) 배경과 고정된 물체는 카메라 프레임 내에서 오른쪽으로 이동되어 보이며, 물체에 의해 생성되는 좌표들은 이동 방향으로 추출될 수 있다.
- 따라서 이동 방향의 흐름을 분석하면, 카메라 이동 인수들의 확장으로 탐지할 수 있으나, 계산하기 위해서는 시간이 소요된다.



(2) 추출된 색상의 표현

- 비디오의 픽셀 데이터는 RGB로부터 Munsell HVC(hue, value, chroma : 색, 명암, 채도) 색상 데이터로 변환된다
- 각 프레임에서 가장 많이 쓰인 색(2-3가지)을 추출하여 대표색으로 만든다. 대표색은 원거리 촬영에서는 배경 장면의 색에 해당되며, 근접 사진이나 7분신 촬영에서는 물체의 색을 나타낸다.

3.3 비디오 색인

비디오 인덱싱 즉, 색인은 비디오 시퀀스에 특정 형태의 표식자를 삽입시키는 과정으로서, 이러한 표식자를 근거로 시퀀스의 검색을 목적으로 하며, 비디오 데이터베이스에서의 내용 기반 액세스는 일반적으로 자동차의 움직임, 걸어가는 사람 등과 같은 특정 시각을 통하는 주요 내용으로 기반한 비디오의 검색을 의미한다.

비디오의 특성을 구체적으로 살펴보면, 비디오란 그 의미와는 직접적인 관계 없이 x, y 축을 가지는 2차원적인 화소열의 시간적인 연속으로 파악될 수 있으며, 비디오 내용 색인이란 비디오와 메타 데이터의 구조화 즉, 비디오 정보 모델 과정과 데이터를 이 모델에 적용시키는 과정으로 볼 수 있다.

비디오의 내용 기반 접근도 위와 같은 방법으로 비디오 데이터를 이미 결정되어있는 어떠한 형태로 구조화 시키고, 구조화된 비디오 데이터를 조직화 시킨다. 비디오에 대한 정보 구조의 설계에 관한 문제들은 비디오 데이터 모델링 분야에서 취급하며, 비디오 처리와 관련된 문헌에서는 비디오 데이터의 적용과 구성에 관한 문제들을 비디오 색인이라 부르기도 한다. 또한 비디오의 색인 방법은 두 가지의 형태로 구분될 수 있는데, High Level Techniques 과 Low Level Techniques의 기술을 들 수 있다. 전자는 미리 정의된 색인 항목들의 집합들을 사용하여 항목화하여금 시간적인 스트림에 따라 비디오를 첨부시키거나 주석을 붙이며, 후자는 이미지나 비디오에 의한 측정값을 따른다. 즉, 비디오로부터 움직임이나 색상과 같은 하위 레벨의 측정치를 추출해 내어 비디오 순서를 위한 키로 사용하는 방법이다. 본 연구에의 비디오 내에서 색인 시스템을 결합하여 비디오를 샷으로 구분하여, 각 샷은 프레임의 표현을 통하여 사용자에게 시각적으로 보여주는 방법에 관하여 연구한다.



가. 색인 형태

비디오 내용 색인의 구분은 시각적이거나 청각적인 대상과 관련한 감각적인 내용과 비디오에 의해 전달되는 함축적인 메시지와 관련된 주제 내용간의 차이점들로 나누거나, 키보드에 의한 기호나 숫자에 따른 색인(정성적인 특징 : qualitative features)과 이미지나 오디오의 특징에 의한 색인(비 가공 특징 : raw features)으로 구분될 수 있다. 여기서 정성적인 특징은 도메인 종속적이며, 도메인 모델로부터 그 특징을 부여 받는다. 이러한 정성적 특징들의 관리를 위해서는 도메인 지식에 대한 풍부한 모델이 필요하며, 또한 유용하다.

비 가공 특징이란 비디오의 컬러 히스토그램이나 이미지 에지 영역과 같은 것으로서, 도메인에는 무관하지만, 함축된 부분의 낮은 단계로부터 색인을 제공한다. 이러한 비 가공 특징들은 감각적인 내용에 한해서 적용될 수 있는 반면, 정성적인 특징들은 감각적이거나 주제적인 내용에 적용될 수 있으며, 이와 같은 색인 형태는 색인을 위해 시공간적으로 확장된 것으로서 어떠한 색인들은 컬러 히스토그램이 매 프레임마다 서로 다른 것처럼 비디오의 단

일 프레임에 한해서 유효할 수 있으며, 비디오 내에서 보이는 사람이나 물체와 관련된 색인에서는 연속된 주기를 가지는 프레임의 내용을 묘사하기도 한다. 이와 유사한 형태로서 프레임의 주기나 프레임 내의 공간적인 영역을 나타내는 색인과 프레임 전체를 나타내는 색인들간의 공간적인 도메인으로 구분 지을 수도 있다.

나. 색인 과정

색인 과정은 크게 수동식과 자동식, 그리고 반 자동식으로 나눌 수 있다.

(1) 수동식 색인 과정

수동식 색인 과정은 키워드 기반 색인으로서 사용자가 수동으로 키워드와 그 내용을 기재하는 것으로 프로덕션 과정이나 포스트-프로덕션 과정의 한 부분으로 이루어진다. 수동 색인 과정의 단점으로는 다음과 같다

- 많은 시간이 소요된다.
- 사용자는 실제 내용의 작은 일부분만 색인 할 수도 있다.
- 색인이 사용자의 주관에 치우칠 수 있다.

위와 같은 단점에도 불구하고, 경험 있는 사서에 의해 색인 과정이 이루어진다면, 색인의 결과는 객관적인 내용을 가지며, 단순한 방법으로 나타낼 수 있다.

다음은 비디오 데이터베이스에 비디오 데이터를 삽입시키기 위해 필요한 과정들을 보인다.

- 이 과정에서는 원래의 비디오 매체를 표준 디지털 형식으로 변환시켜 디지털 비디오화 된 것으로 가정한다.
- 비디오 조작 명령 : play, stop, rewind, pause, fast, forward, still, step, forward, step reverse 기타.
- 다음 과정은 비디오 데이터 모델로 비디오를 입력시키는 과정을 나타낸다.

단계 ① : Approximate shot boundary location : 비디오 재생 → 첫 번째 샷 경계를 정하

면서 → still, stop 등의 제어를 통해 경계 부분의 근접한 위치를 표시한다.

단계 ② : Exact boundary location : 경계의 프레임 숫자로서 샷 경계의 세밀한 위치를 잡는다 이 과정에는 확실한 경계를 잡기 위해 frame step 기능을 사용한 검색 과정이 포함된다.

단계 ③ : Shot Viewing : 주변과 분리된 샷을 보는 과정이다. 이 과정은 샷의 인식이 샷의 내용에 영향을 받지 않는 것이 중요하다.

단계 ④ : Shot Feature Evaluation : 이 과정에서는 오퍼레이터가 시각적으로 모델에서 다양한 특징들을 추출해내는 것으로 영화적 요소에 대한 내용을 잘 알고 있어야 한다. 일반적으로 higher level labels이 사용되며, 오퍼레이터는 비디오의 영역에 대해 잘 알고 있어야 한다.

단계 ⑤ : Shot Feature Logging : 이 단계에서는 추출된 특징들을 데이터 모델 구조로 저장시키는 과정으로서 단계 4, 5는 마우스, 키보드 등을 사용한다.

· 위와 같은 일련의 과정들은 비디오 데이터들을 수동으로 비디오 데이터 모델에 입력시키는데 필요한 과정이다.

· 이러한 과정을 평가하기 위해서는 Insertion Time, Insertion Effort, Insertion Errors와 같은 3가지 고려 사항이 있다. 이러한 요소들은 총 비용 계산에 있어서도 매우 중요하다

(2) 자동식 색인 과정

자동식 색인 과정은 수동식 색인 방식에 따른 비용 문제 해결을 위한 것으로 자동식 색인 방식을 사용한다. 이 과정은 컴퓨터 제어에 의한 이미지나 비디오의 분석 방법으로서 low-level, intermediate-level, high-level 처리의 3단계로 나눌 수 있다.

① low-level processing : 분석을 위해 가능한 한 좋은 데이터 기반을 가지기 위해 노이즈 감쇄와 같은 이미지를 향상시키는 처리 단계

② Intermediate-level processing : 이 단계에서는 이미지를 분할시켜 영역과 같은 주요 성분들을 추출한다.

③ High-level processing : 이 단계는 이미지 내의 대상 물체를 인식하여 그 이미지의 내용을 해석하기 위한 목적이다. 비디오의 장면 분석은 이미지 분석과 유사한 것으로서 비디오에서는 움직임 정보를 이용할 수 있으므로 세그멘테이션과 같은 처리 과정을 더 쉽게 할 수

있다.

자동식 과정의 장점으로는 객관적이며, 사용자가 원래 색인되지 않은 대상들의 검색을 원하는 경우에도 사용할 수 있다. 자동 색인 과정에서는 문제점은 특히 상위 계층의 처리 과정에서 나타난다. 컴퓨터에 의한 이미지 해석은 매우 복잡한 과정으로서, 처리되어야 하는 방대한 데이터량과 자세한 이미지 내용과 같은 원하는 결과로부터 얻을 수 있는 기본적인 처리 도구들이 부족하기 때문에 많은 어려움이 야기된다.

다. 색인 방식

(1) Segmented indexing

종래의 비디오 색인 방법은 비디오를 chunks로 세그먼트시키는 것으로, 키 워드의 집합은 비디오의 각 chunk와 관련되며, chunk의 내용을 색인한다. 세그멘테이션 방식의 장점으로는 주어진 chunk에 대한 모든 색인을 찾는 것이 일괄적인 반면, 모든 키워드가 비디오 chunk와 관련되기 때문에 중요한 문제를 야기시킬 수도 있다.



(2) Stratification

세그멘테이션에 의한 문제를 극복하기 위한 방법으로서 층을 이루게 하는 개념인 계층화 방식이 있다. 이 개념은 비디오를 나누는 대신 그 기술 내용을 나누는 개념이다. 즉, 각 키워드는 구분된 비디오 단편들과 관련이 있다. 이와 같은 방법이 가지는 단점으로는 비디오 프레임의 주어진 주기와 연관된 다양한 색인을 찾기 위해서는 많은 계산 과정이 요구된다.

(3) 계층적 색인(Hierarchical indexing)

계층적 방법은 개별 색인들을 결합시켜 더 복잡한 개념으로 만들게 하는 것으로서 대수 비디오 시스템에서 비디오 편집을 위해 사용하는 대수 연산을 정의하고 있다. 대수 연산은 시스템에서 연결된 계층들을 정의하는데 사용되고 있으며 계층들을 연결시킴으로써, 시스템은 같은 비디오 단편과 연관된 서로 다른 계층들간의 연관 관계를 규정하는데 있어서 원래의 계층화 접근 방식보다 더 직접적인 방법을 제공해준다.

IV. 내용기반 정보 알고리즘

4.1 내용기반 검색 알고리즘

본 절에서의 내용기반 검색은 색상을 이용한 특정 물체를 추출해 낼 수 있는 방법에 대해 기술한다. 색에 의한 이미지 정보 검색은 색상 히스토그램 중복도, 색상 히스토그램 거리 가중치, 평균 색상 거리, 색상 인정 정보 등을 이용한 색상 유사도를 측정함으로써 이루어진다. 히스토그램 기반의 방법은 장면 변화를 검출하기 위해 가장 많이 쓰이고 있는 방법으로, 같은 장면으로 분류해야 할 프레임들의 색상 분포는 거의 비슷하다는 속성을 사용하는 방법이다. 이러한 색에 의한 정보 검색에서 문제시 되어 왔던 지역 칼라 영역의 자동 추출, 효율적인 인덱싱, 그리고 효과적인 검색의 동시 만족 요구를 이진 칼라 집합의 역투상을 사용하여 해결하고 있다. 이진 칼라 집합의 역투상을 구별되는 칼라 영역을 규정하기 위한 것으로, 방대한 수의 칼라 집합에 반복 적용함으로써 지역적이고 임의의 형태를 지닌 칼라 영역을 추출해낸다. 또한 하나의 영역으로부터 추출된 크기나 위치, 모양과 같은 정보도 이미지 데이터를 검색하는데 이용된다. 색상을 이용한 검색 방법으로 본 논문에서 제안하는 방법은 색상 컬러 특징을 이용한 추출과 영상의 컬러, 밝기, 에지 성분을 가지고 영상의 중요한 특징을 효과적으로 표현할 수 있는 히스토그램의 생성법등 2가지 방법을 제안하며, 이러한 알고리즘인 색상을 이용해 Video 데이터 상에서의 특정 물체를 추출해 낼 수 있는 색상 정보와 색상 변환처리를 제안한다

제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

가. 색상 Color 특징 추출

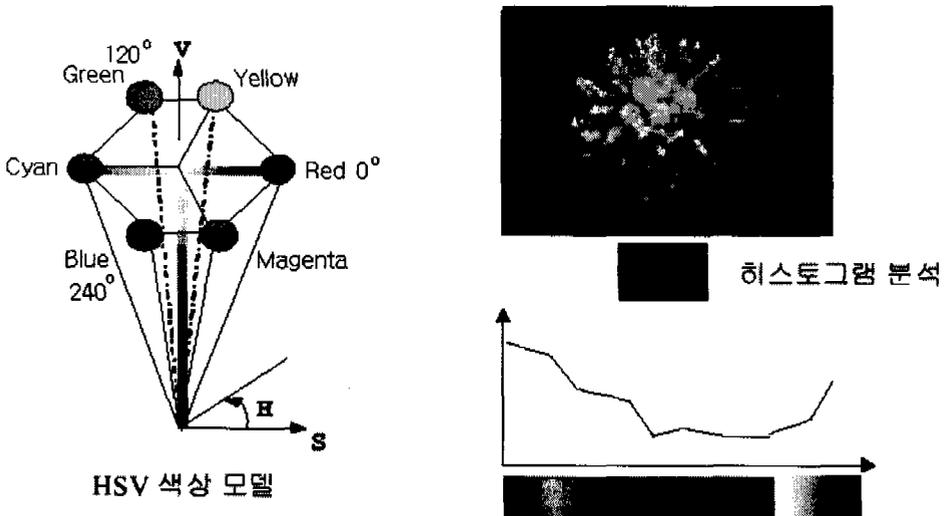


Fig. 8 HSV 색상 히스토그램

색상 기반 검색 모듈은 색상 히스토그램 기법을 사용한 영상 검색 방법의 하나로 영상에서의 객체의 위치나 회전 그리고 어느 정도의 조명 변화에 견고한 특징을 가지고 있다. 본 논문에서는 인간의 시각 특징과 유사하다고 알려진 HSV 색상 모형하에서 색상 히스토그램 분석을 통해 영상을 색인하는 알고리즘을 모듈화한다.

히스토그램 분석을 위한 각 히스토그램 빈의 구성은 다음 표와 같다

Bin Number	Hue	Value	Saturation
1	0~13 350~359	30~100	40~70
2	0~13 350~359	30~100	70~100
3	13~25	70~100	70~100
4	40~68	60~100	60~100
5	13~25	40~100	30~65
6	25~45 345~359	30~80	30~100
7	50~165	24~100	20~100
8	165~230	50~100	25~100
9	230~267	30~100	35~100
10	257~308	40~100	25~100
11	308~359	40~100	30~100
12	-	0~20	-
13	-	60~80	0~5
14	-	20~50	-
15	-	90~100	0~20

Table. 1 히스토그램 분석을 위한 각 히스토그램 빈의 구성표

나. Composite(Color/Edge) 특징 추출기

본 절에서 제안된 영상의 컬러, 밝기, 에지 성분을 가지고 영상의 중요한 특징을 효과적으로 표현할 수 있는 히스토그램의 생성법은 기존의 방법보다 적은 수의 빈으로 보다 효과적인 영상 검색을 할 수 있는 영상의 표현자(Descriptor)가 될 수 있다. 여기에는 히스토그램의

각 빈에 대한 의미적 정의를 최적화 함으로써 적은 수의 히스토그램 빈으로 영상의 특징을 잘 반영할 수 있다. 또한 적은 수의 히스토그램 빈으로 계층적 구조를 갖는 히스토그램을 만들어 영상 내 스케일러빌리티(Scalability)를 들 수 있다.

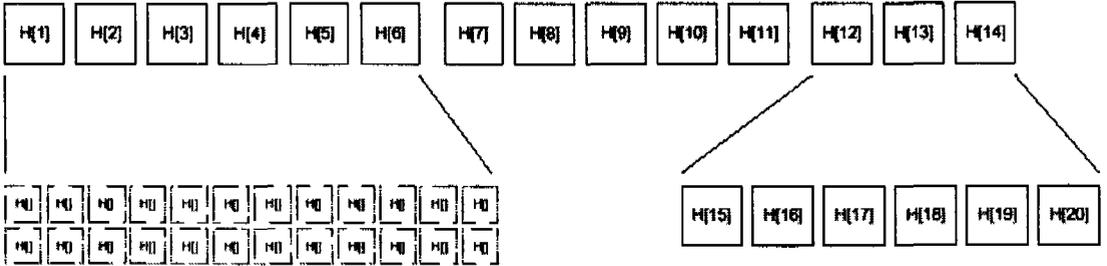


Fig.9 계층적 구조의 히스토그램

제안된 히스토그램은 Fig. 9 와 같이 최상위 레벨에서 작은 수의 기본 빈들로 구성된 계층적 구조를 가지고 있다. Table. 2 는 최상위 레벨의 각 빈들이 갖는 의미를 나타낸 것으로 컬러, 밝기, 에지 정보들에 대하여 기본 빈들을 표현한 것이다. 최상위 레벨의 빈들을 히스토그램의 형태로 표현한 것이다. 여기서 볼 때 하위 레벨로 내려감에 따라 빈의 개수를 늘여서 Table. 3 과 같이 다양한 정보를 사용하여 영상을 보다 정확하게 표현 할 수 있다.

Index	Classification	Description	Index	Classification	Description
H[1]	M	Pure color magenta	H[12]	V	Vertical Edge
H[2]	B	Pure color blue	H[13]	H	Horizontal edge
H[3]	C	Pure color cyan	H[14]	Mix	Complex edge
H[4]	G	Pure color green			
H[5]	Y	Pure color yellow			
H[6]	R	Pure color red			
H[7]	L0	Black			
H[8]	L1	Dark Gray			
H[9]	L2	Gray			
H[10]	L3	Light Gray			
H[11]	L4	White			

Table 2 최상위 레벨에서 기본 빈의 의미 - (1)

Index	Classification	Description	Index	Classification	Description
H[15]	Me	Edge color magenta			
H[16]	Be	Edge color blue			
H[17]	Ce	Edge color cyan			
H[18]	Ge	Edge color green			
H[19]	Ye	Edge color yellow			
H[20]	Re	Edge color red			

Table. 2 최상위 레벨에서 기본 bin의 의미 - (2)

Table. 2 에서 나타난 에지 정보의 추출은 기존의 전통적인 에지 검출 방법 등 다양한 방법을 사용할 수 있다. Table. 3 과 같이 슈퍼픽셀(superpixel)에서 에지를 추출하는 방법에서 슈퍼픽셀은 영상 내 픽셀들의 집합을 의미한다. 슈퍼픽셀에서 에지를 추출하는 방법은 슈퍼픽셀에서 4개의 부-슈퍼픽셀로 나누어 각 부-슈퍼픽셀 내 평균 밝기 값의 차이를 사용하여 부-슈퍼픽셀간 에지를 검출하게 된다. 그리고 에지의 카운트 방법을 기존의 선형방법을 사용하지 않고 개발된 알고리즘을 통한 연결된 에지를 위한 가중치를 사용하여 적용한다.

$$H_{edge} = \sum_{i=1}^M i + \sum_{i=1}^M i + \dots \dots \dots \quad (\text{식 3-1})$$

일반적으로 색상은 약 400nm에서 700nm까지의 파장을 지닌 가시광선 영역의 전자기파를 의미하는 것이다. 모든 색상은 소위 주요 컬러(RGB)라는 서로 다른 파장의 길이를 가지는 세 가지 성분의 조합과 빛의 휘도 성분으로 표현된다. 인간의 망막으로 받아들이는 위와 같은 세 가지의 주요 컬러 성분은 서로 독립적인 컬러 매칭 함수와 적절한 스펙트럼 함수로서 표현된다.

$$R = \int_{v_1}^{v_2} r(v)U(v)dv \quad , \quad G = \int_{v_1}^{v_2} g(v)U(v)dv \quad , \quad B = \int_{v_1}^{v_2} b(v)U(v)dv \quad (\text{식 3-2})$$

여기서 v 는 빛의 주파수이다. 대부분 사용되는 하드웨어의 종류에 따라 여러 가지 컬러 모델을 사용한다. 일반적으로 비디오 카메라는 RGB모델을 가장 많이 사용하고 있으며 이 모델은 다른 컬러 모델로 쉽게 변환이 가능하다. 비디오 카메라나 프레임 그래버는 아날로그 신호를 각각의 화소에 대해 세 개의 [0,1,...,255]의 값으로 변환한다. 즉, 한 화소는 각각

[0,1,...,255]값을 가지는 RGB의 조합 $Q = (R, G, B)$ 으로 표현된다. 하지만 이 모델은 하나의 색상뿐만 아니라 빛의 휘도 까지 포함된 값이다. RGB공간상에서의 한 점 $Q = (R, G, B)$ 는 색상뿐만 아니라 휘도를 지니고 있다. 만약, 공간의 두 점 $Q_1 = [r_1, g_1, b_1]$ 과

$Q_2 = [r_2, g_2, b_2]$ 가 (식 3-3)와 같은 관계를 지니고 있다면,

$$\frac{r_1}{r_2} = \frac{g_1}{g_2} = \frac{b_1}{b_2} \quad (\text{식 3-3})$$

이 두 점은 같은 색상을 가지고 있으나 다른 휘도값을 가지고 있다는 것을 의미한다. 인간의 시각 체계는 여러 가지 다른 밝기와 환경에 의한 조명의 변화에 대해 색상을 구분할 수 있는 능력을 지니고 있으며 순수한 하나의 색상을 표현할 때 휘도는 적은 영향을 미친다. 그리고, 조명의 변화에 색상은 영향을 받지 않으나 휘도는 큰 변화를 가지게된다. 따라서, RGB 공간상의 한 점에 대해 정확한 색상을 유지하면서 조명의 변화를 제거할 수 있는 정규화 과정을 거칠 필요가 있다.



$$\begin{aligned} r &= \frac{R}{R+G+B} \\ g &= \frac{G}{R+G+B} \\ b &= \frac{B}{R+G+B} \end{aligned} \quad (\text{식 3-4})$$

RGB 모델의 3차원 공간상의 점들이 위와 같은 정규화과정을 거치게 되면 순수한 컬러 정보만을 지니게 된다.

$$r + g + b = 1 \quad (\text{식 3-5})$$

그리고 (식 3-5)와 같은 관계를 가지게 되어 세 개의 요소 중 두 개의 요소로 컬러를 표현할 수 있다. 즉 $R^3 \Rightarrow R^2$ 의 매핑 관계로 차원을 낮출 수 있어 계산상에서도 효율적이다. 정규화 과정을 거친 색상 모델에서는 화소에 대한 색상 정보를 $q = (r, g)$ 로 나타낼 수 있으

며, 이는 각 색상 성분이 8비트로 나타내 진다면 기존 RGB 모델에서 표현 가능한 256^2 가지의 색상 정보중 휘도의 변화에 덜 민감한 $255^2/2$ 개의 색상정보를 정규화 색상 모델에서 표현할 수 있음을 의미한다. 또한 이 정규화 과정은 인간의 시각체계를 잘 표현한다는 HSI 모델과 밀접한 관련을 지고 있다.

$$\begin{aligned}
 H &= \cos^{-1} \left\{ \frac{\frac{1}{2} [(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right\} \\
 S &= 1 - \frac{3}{(R+G+B)} [\min(R, G, B)] \\
 I &= \frac{1}{3} (R+G+B)
 \end{aligned}
 \tag{식 3-6}$$

즉, 정규화 과정을 거친 RGB 모델은 HSI 모델에서 밝기 성분을 제거하고 Hue와 Saturation 성분을 구 좌표계에서 직교 좌표계로 나타냈다고 할 수 있다. 위의 식에서도 알 수 있듯이 HSI 모델보다는 정규화를 거친 RGB 모델이 더 수식적으로 간단하고 계산상에서도 효율적이다.



4.2 MB 기반 컷 검출 알고리즘

본 절 부터는 MPEG-2 Video 상에서 압축된 영역의 특징물을 추출하는데 있으며, MPEG으로 압축된 비디오에서 키 프레임을 정의하는데 어려움은 프레임들이 서로 다른 종류(즉, I, P, B 프레임)라는 것과 다양한 패턴으로 나타난다는 것이다. 본 연구의 실험은 카메라를 통하여 저장되어진 비디오 데이터를 기반으로 실험한 내용에 대하여 기술하였고, 비디오 데이터 상에서의 이미지 검색은 두 가지의 방법으로 실험을 하였다. 첫 번째로, 사용자로부터 스캐치된 이미지를 입력받아 이미지 데이터베이스로부터 추출된 에지와 세션화된 이미지와의 관련성에 의해 검색을 수행할 수 있는 모양에 의한 이미지 검색 방법과, 두 번째로 이미지 쌍들의 공통적인 특징들을 적절히 추출하여 사용하는데 기반하며, 이미지들은 n-차원의 특정한 방향의 선들과 n-차원 공간에서의 점들과 같은 내용에 의해 표현되는 유사측정을 통한 이미지 검색방법에 대해 기술한다.

본 실험에서는 프레임 형태에 독립적인 표현이 필요하게 된다. 이 표현에서 인덱싱과 검색 단계를 위해 추출한 모든 특징들은 프레임 형태와 같은 요소들과는 독립적으로 얻어지게

된다. 또한 본 실험에서는 인덱싱과 검색 단계에서, 키 프레임의 DC 계수와 움직임벡터를 이용한다. MPEG에서 MB는 MB가 인트라 부호화된 것인지, 순방향/역방향으로 예측된 것인지, 혹은 쌍방향으로 예측된 것인지에 상관없이 각각의 프레임 형태에 따라 0, 1, 2개의 움직임벡터들을 가질 수 있다. 더욱이, 이런 움직임벡터들은 그 MB가 속해있는 프레임에 대한 레퍼런스 프레임들에 대해 순방향으로 예측되거나 역방향으로 예측될 수 있다. 만약, 움직임벡터가 없는 I 프레임과 쌍방향으로 예측된 MB들을 갖는 B 프레임을 비교할 때, 혹은 하나는 순방향으로 예측되고 다른 하나는 역방향으로 예측된 두 개의 B 프레임들을 서로 비교할 때 문제가 발생한다.

비디오의 속성을 보면 I와 P 프레임만을 갖는다. I 프레임들 사이에 P 프레임들만 있다면 그리고 B 프레임들이 없다면, 흐름은 두 개의 연속되는 I 프레임들 사이에서 각 프레임들에 대해 유도될 수 있다. 여기서 I 프레임들은 자신을 포함하고 마지막 P 프레임은 제외한다. 마지막 P 프레임을 제외하는 이유는 마지막 P 프레임 다음에 오는 I 프레임과의 관계에 대한 정보를 아무것도 가지고 있지 않기 때문이다. P 프레임이 뒤 따르는 I 프레임에 대한 흐름은 순방향으로 예측된 P 프레임을 역(inversion)을 취한 후의 움직임벡터들의 집합이다.

반면에 대부분의 MPEG 스트림은 연속되는 레퍼런스 프레임들 사이에 한 개 이상의 B 프레임들을 갖는다. 두 개의 연속되는 레퍼런스 프레임들 R_1 와 R_2 로 놓자. 이 레퍼런스 프레임들 사이에 B 프레임들을 B_1, \dots, B_n 으로 나타내고, 여기서 n 은 이 두 레퍼런스 프레임들 사이의 B 프레임의 수를 의미한다(일반적으로, $n = 2$). 첫 번째 단계는 B_1 의 순방향으로 예측된 움직임벡터들을 이용하여, 첫 번째 레퍼런스 프레임 R_1 와 그 다음 프레임 B_1 사이의 흐름을 유도하는 것이다. 이 경우는 위에서 설명한 I-P 경우와 비슷하다. 순방향으로 예측된 움직임벡터들의 역(inverse)은 R_1 의 MB들에 대한 흐름벡터(flow vector)를 형성한다. 비슷하게, R_2 에 대한 프레임 B_n 의 역방향 움직임벡터를 이용하여 프레임 B_n 에 대해 흐름이 유도될 수 있다. 여기서, 움직임벡터들의 역을 취할 필요는 없다 왜냐하면 이런 경우의 흐름벡터들은 기본적으로 역방향으로 예측된 벡터들이기 때문이다. R_2 가 그것을 뒤따르는 레퍼런스 프레임을 이용하여 분석되어질 때 R_2 에 대한 흐름이 유도된다.

매크로블록 기반 컷 검출 알고리즘은 각 프레임의 MB들의 형태를 이용하여 컷 검출 여부를 결정한다 Table. 4은 이 장에서 이용된 변수들을 정의한 것이다. 여기서 이용된 임계값 t_1, t_2, t_3 는 MB들을 이용한 컷 검출 과정에 이용된다. 그리고 t_4, t_5 는 DCT 비교법(DCT

comparison)에 적용된다. MB들의 총 갯수와 스킵된 MB들의 갯수를 식 (3-7)과 같이 정의한다.

$$\begin{aligned} N &= N_f + N_b + N_{fb} + N_s \\ N_s &= N_{sf} + N_{sb} + N_{sfb} \end{aligned} \quad (\text{식 3-7})$$

변수 정의	
N	한 프레임에서의 총 MB들의 수
N_f	순방향으로 예측된 MB들의 수
N_b	역방향으로 예측된 MB들의 수
N_{fb}	쌍방향으로 예측된 MB들의 수
N_s	스킵된 MB들의 수
N_{sf}	순방향으로 예측되어 스킵된 MB들의 수
N_{sb}	역방향으로 예측되어 스킵된 MB들의 수
N_{sfb}	쌍방향으로 예측되어 스킵된 MB들의 수

Table 4. MB 변수 정의

클립에서의 I 프레임들의 집합을 I, P 프레임들의 집합을 P, B 프레임들의 집합을 B로 놓자. SV_i 는 수식 3-8와 같이 비디오 클립에서 각 프레임 i 에 지정된다.

만약 비디오 클립이 I와 P 프레임만을 갖는다면, 프레임 i 에서 시작하는 새로운 시퀀스가 결정되어진다($i \in P$, if $SV_i/N < t_1$).

$$SV_i = \begin{cases} 0 & \text{if } i=0 \\ SV_{i-1} & \text{if } i \in I \\ N_f + N_s & \text{if } i \in P \\ N_{fb} + N_{sfb} + N_f + N_{sf} & \text{if } i \in B \end{cases} \quad (\text{식 3-8})$$

연속되는 I 프레임들의 DCT 분석은 I 프레임전에 발생하는 샷변화를 검출하기 위해 수행된다. DCT 분석에 의해 연속되는 I 프레임들 사이에서 컷이 검출되고 I 프레임들 사이의 P 프레임의 MB들의 형태 대부분이 순방향 예측되었다면, 후자의 I 프레임에서 시작하는 새로운 시퀀스가 결정되어진다.

한 클립에서 P 프레임들이 존재한다면, 일반적으로 B 프레임들 또한 존재한다. 비디오 클립이 B 프레임들을 갖는다면, R을 레퍼런스 프레임들의 집합으로 놓자, $R = I \cup P$. 어떤 두 레퍼런스 프레임들 R_i 와 R_j 사이에 후자의 레퍼런스 프레임 R_j 와 함께 모든 B 프레임들을 포함하는 집합을 domain D로 정의하자.

$$R, BB \dots BR, (\text{domain } \mathbb{D} = BB \dots BR) \quad (\text{식 3-9})$$

어떤 프레임 $k \in \mathbb{D}$ 에 대해, 만약 Table. 5에 서술된 네 가지 조건들을 만족한다면, 프레임 k 에서 시작하는 새로운 시퀀스가 결정되어진다.

	조 건
조건 1	만약 $R_i \in P$ 이면, $SV_i/N < t_2$
조건 2	$\forall l \in \mathbb{D}$ 일 때, $SV_l/N < t_3 (l \neq B)$
조건 3	$\forall \mathbb{D}$ 에서 k 이전의 B frames, $N_f \cdot N_b > N_{fb}$
조건 4	$\forall \mathbb{D}$ 에서 k 이후의 B frames, $N_b \cdot N_f > N_{fb}$

Table 5 MB 기반의 컷 검출 조건

만약 조건 1은 해당 프레임의 검사여부 결정에 이용된다. 만약 이 조건이 거짓이면, 남은 조건들에 대한 테스트는 모두 생략된다.

4.3 특징들을 이용한 비디오 인덱싱

본 실험의 과정인 압축된 영역에서의 특징추출 기법과 움직임벡터를 이용하여 사용자가 관심을 갖게되는 비디오 데이터를 구조화하고 검색할 수 있는 인덱싱 기법을 연구한다

Fig 10 은 MPEG 비디오에서 컷 검출에 의해 분해된 샷을 인덱싱하는 과정을 간략하게 보여준다.

MPEG에서 일련의 비디오클립이 주어지면, 질의를 통해 비디오 클립에 대한 내용을 검색한다. 이러한 질의들이 가능하도록 비디오클립들을 구조화해야 한다. 비디오클립은 공간적 차원과 시간적 차원 모두를 가지고 있으므로, 본 논문에서는 각 장면들의 키 프레임은 공간적 차원을 표현하는 그 DC 계수들과 시간적 차원을 표현하는 그 흐름벡터들과 함께 그 장면에 대한 인덱스로서 받아들인다. 만약 질의가 단 한 프레임이라면, DB에 저장된 비디오들의 움직임정보와 대조하기 위해 끌어내야 하는 시간적 정보는 필요 없게 된다.

두 프레임들 사이의 공간적 유사성은 프레임들이 휘도, 색도, 질감, 형상과 같은 비슷한 공간적 특성들을 갖는다는 것을 의미한다. 공간적 유사성을 찾는데 가장 간단한 접근방법은

모든 클립의 매 프레임에 대한 모든 DCT 정보를 저장하고 비교 검색하는 것이다. 질의 프레임이 주어질 때, 가장 비슷한 한 프레임을 찾기 위해 가능한 한 모든 정보와 비교해야 할 것이다. 그러나, 이 접근방법은 시간과 공간모두에서 비효율적이다.

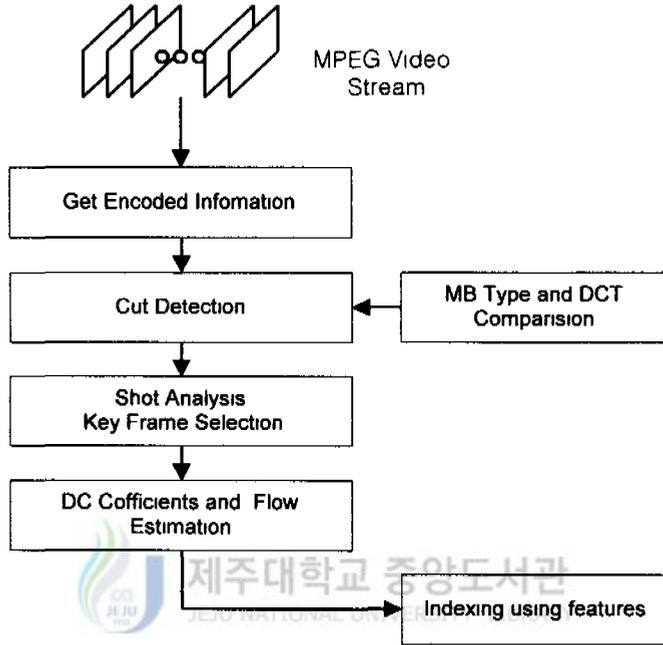


Fig. 10 MPEG-2 Video의 video index의 흐름

그리고, 휘도와 색도 성분의 DC 계수만을 이용하고 공간적 유사성을 분석하기 위해 그 계수들과 다른 프레임에서 부합하는 DC 계수들을 비교한다. 각각의 프레임을 DC 계수들의 벡터로서 다루었고 프레임들의 유사성을 결정하기 위해 벡터들사이의 거리함수를 위한 값을 이용한다.

V. 실험 및 결과 고찰

MPEG 스트림에서 간단하고 효과적인 인덱싱 방법을 제공하기 위해 본 논문에서는 공간과 시간적 유사성 기법을 결합하였다. 먼저, 각 MB마다 6개의 DC 계수들로부터 휘도성분의 평균치와 색도성분의 평균치, 그리고 각 DC 계수들을 적용된 B^+ 트리에서 키 프레임으로 인덱스한다. 질의가 단 하나의 프레임으로 구성되어 진다면 질의에 가장 비슷한 키 프레임을 찾기 위해 DB에 저장된 휘도성분과 색도성분의 평균치를 비교하고, 그 비교치가 임계값에 이르면 정확한 매칭을 위해 각 DC 계수들을 비교한다. 이때는 움직임벡터를 추정할 필요는 없다.

비디오 검색에 있어서는 질의 키 프레임과 처음으로 가장 비슷한 몇 개의 프레임들을 돌려줄 수 있고 그것을 사용자 브라우저 결과로 생각할 수 있다. 짧은 질의 시퀀스의 경우에 질의 시퀀스의 처음 몇 프레임들에 대해 질의 하고, 질의에 적당한 것들을 구별하기 위해 테이블에 저장한다, 이런 키 프레임들은 대표 프레임으로 다루어지고, 그들의 순위를 변경하기 위해 움직임 정보들을 비교한다. 실험을 위해, 본 논문에서는 5~30 fps(frame per second)으로 변화하는 프레임비율(frame rate)로 디지털화한 대략 748 프레임들을 갖는 총 8 개의 MPEG 비디오를 이용하였다. 본 실험에서는 Visual C++ 환경에서 수행되었다.

가. MPEG에서의 컷 검출 실험

MPEG 스트림에서 컷은 대부분 다음과 같은 세 가지 경우에 발생한다.

- ① 어떤 P 프레임이 인트라 부호화된 MB들이 대다수일 때
- ② B 프레임에서 대다수의 MB들이 이전 I나 P 프레임들로부터 순방향으로 예측되었을 때
- ③ B 프레임에서 대다수의 MB들이 역 방향으로 예측되었을 때.

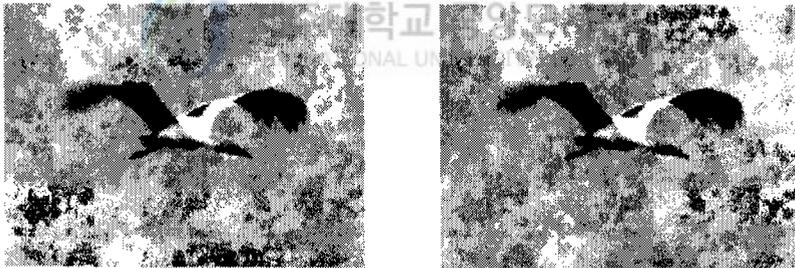
Table. 6은 이러한 단계의 프레임에서 컷의 발생된 프레임에 따른 조건 값들을 보여준다. 여기서 B_2 , I, B_1 Frame에서 B_2 Frame에서 컷이 검출된다.

	SV _i /N	SV _i /N	N _{r(i-1)} -N _{b(i-1)}	N _{fb(i-1)}	N _{b(i)} -N _{r(i)}	N _{fb(i)}
B ₂ Frame	Passed	Passed	27	317	16	307
I Frame	Passed	1.77 (>t ₃)	Passed		Passed	
B ₁ Frame	Passed	Passed	Passed		-374	8
B ₂ Frame	Passed	Passed	374	8	395	1

Table. 6 Result of condition values

나. 비디오 질의 및 검색

DB에 저장된 키 프레임의 DC 계수들과 움직임 정보를 비교하고 질의 키 프레임의 DC 계수들과 움직임 정보를 비교하여 비슷하게 매치(match)하는 시퀀스들을 돌려준다. 검색 시간을 줄이기 위해, 먼저 적용된 B'트리의 각 노드에 저장된 DC 계수들과 움직임벡터의 평균치를 비교한 후, 비슷하게 일치하면 더욱 정확한 질의결과를 산출하기 위해 그 샷을 검색하였다. Fig. 11 은 하나의 질의 프레임이 주어졌을 때, DB에 인덱스된 정보와 추출된 질의 프레임의 특징들을 비교하여 비슷한 프레임들을 돌려주는 것을 보여준다.

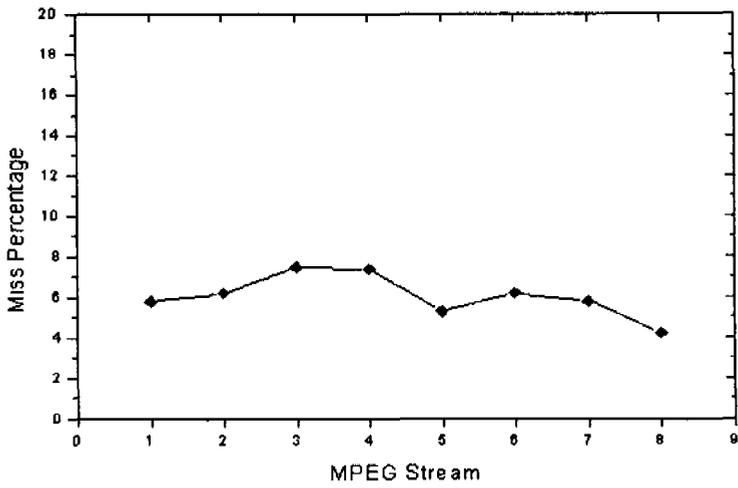


Query Frame

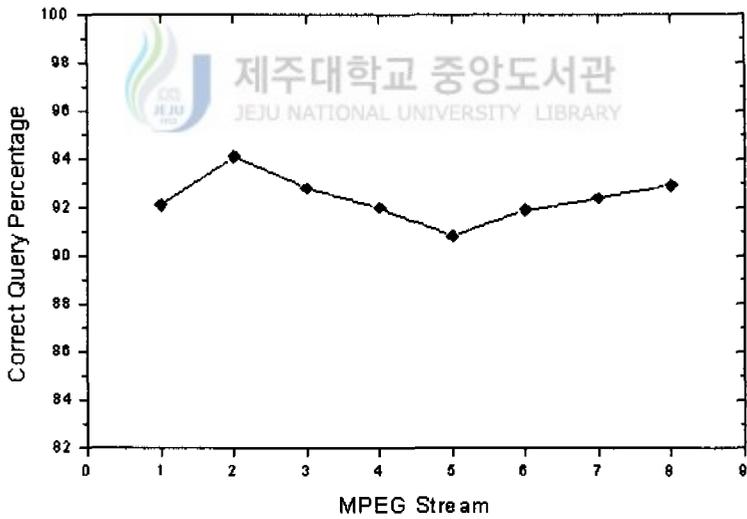
Match 1(Frame 212)

Fig. 11 Retrieval results on MPEG Video Stream A (Continue)

이러한 기법들을 이용하여 픽셀영역과 압축된 영역에서 특징 추출 후에 인덱싱하는 과정까지의 소요된 시간을 보여주고 있다. 픽셀영역에서 인덱싱하는 것은 MPEG 비디오를 압축 해제한 후에 그 실제적인 처리가 이루어지고 그 처리단위가 픽셀단위이므로 그 만큼 많은 시간이 소요되었다.



(a) Query Frame Percentage



(b) Match frame Percentage

Fig. 12 Percentages of retrieval results on MPEG stream

VI. 결 론

본 논문은 멀티미디어 정보를 고품질 고음질로 압축 및 복원하는 MPEG-2 압축 영상을 효과적으로 저장하고 검색하는 압축 동영상 내용에 기반한 정보검색을 연구하였다.

내용기반 멀티미디어 정보검색은 일반적인 정보를 캡션(Caption) 정보에 기반하여 정보검색 기능을 수행할 수 있을 뿐 아니라, 멀티미디어 정보를 내용에 따라 사용자가 원하는 데이터를 제공할 수 있는 기능을 지원한다. 본 논문에서는 동영상 저장 및 검색 시스템은 멀티미디어 정보검색 시스템의 하나의 독립적인 모듈로서, 동영상의 내용에 기반하여 사용자가 원하는 영상 정보를 수 초 혹은 수 분의 연속 장면, 또는 전체 장면들 각각을 독립적인 정보의 단위로 취급하여 저장 및 검색 할 수 있는 기능을 제공하였다.

컷검출 기법과 DC 계수와 흐름정보 추출기법을 이용하여 인덱싱을 수행한 결과 90% 이상의 질의 정확성을 보였다. 이는 압축된 영역에서 압축해제없이 부호화된 정보만을 이용한 본 기법이 압축된 영역에서 효과적이라는 것을 입증한다, 이러한 기법들을 이용하여 대용량의 영상 데이터베이스에 적용할 수 있다.

향후 연구 과제로는 기존방법들과의 비교를 통해 정확하고 효율적인 검색방법을 모색하고 다양한 영상의 검색에 응용하는 연구가 필요하다.

參 考 文 獻

1. F. Arman, A. Hsu, and M.Y. Chiu, "Image Processing on Compressed Data for Large Video Databases", Proc. ACM Multimedia 93, Anaheim, CA, pp. 267-272, 1993.
2. D. Le Gall, "MPEG: A Video Compression Standard for Multimedia Applications", Communications of the ACM, 34(4), pp. 46-58, 1991.
3. R. M. Haralick and L. G. Shapiro, "Computer and Robot Vision", Vol. 2, Addison-Wisely, Reading, MA, 1993.
4. M. Mills, J. Cohen, and Y. Y. Wong, "A Magnifier Tool for Video Data", Proc. CHI '92, Monterey, CA, pp. 93-98, 1992.
5. A. Nagasaka and Y. Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances", Visual Database Systems II, E. Knuth and L. M. Wegner, editors, North-Holland, pp. 119-133, 1991.
6. B. C. O'Connor, "Selecting Key Frames of Moving Image Documents: A Digital Environment for Analysis and Navigation", Microcomputers for Information Management 8(2), pp. 119-133, 1991.
7. R. Steinmetz, "Data Compression in Multimedia Computing-Standards Systems", Multimedia Systems 1(4), pp. 187-204, 1994.
8. D. Swanberg, C. F. Shu, and R. Jain, "Knowledge Guided Parsing in Video Databases", Proc. IS&T/SPIE Conf. on Storage and Retrieval for Image Video Databases, San Jose, CA, 1993.

감사의 글

먼발치로만 봐도 좋고, 가까이 다가가 옆에 있으면 더 좋고, 저항할수 없는 이상한 힘에 이끌려가는 느낌이지만 거기엔 기쁨과 고마움이 있습니다. 항상 부족하기만 한 저이기에 제가 받은 사랑과 도움에 대하여 감사드릴 분들이 너무나 많습니다.

지금의 제가 있기까지 사랑과 희생으로 보살펴 주신 부모님의 은혜에 감사드리며 이 조그만 학문의 결실을 드리려합니다. 또한 멀리서 지켜보시는 장인, 장모님께서는 미안함과 고마움을 전하며 앞으로 더 나은 모습을 보여드릴 것을 약속 드립니다.

어려운 시기에 물심 양면으로 도움을 주신 큰아버님, 큰어머님, 고모님, 수근형, 형수님 은혜도 가슴속 깊이 새기겠습니다.

서툴고 미숙했던 지난 석사과정동안 관심과 사랑으로 지켜봐 주시고 자상함과 따뜻함으로 지도해 주신 김장형 교수님, 본 논문이 완성되기까지 세심한 지도와 격려 인간적인 가르침을 주신 안기중 교수님, 보다 나은 논문이 되도록 세심하게 지도하여 주신 이상준 교수님, 대학원 생활동안 많은 가르침을 주신 곽호영 교수님, 변상용 교수님, 송왕철 교수님, 이동희 교수님께 머리 숙여 감사드립니다.

관심과 사랑으로 지켜봐 주신 제주산업정보대의 이봉헌 교수님, 홍석표 교수님, 고석만 교수님, 김근형 교수님, 박창희 교수님, 이철호 교수님, 이은주 교수님께도 감사드립니다. 그리고 대학원 생활 동안 곁에서 정신적, 학문적으로 큰 힘이 되어 주었던 진석, 무영, 영수, 동진,길봉형, 지환형에게도 깊은 사랑의 마음을 전합니다. 마지막으로 따뜻한 사랑을 준 형제들에게도 깊은 사랑의 고마움을 전합니다.

어려움 속에서도 사랑으로 뒷바라지를 해준 사랑하는 아내 진옥, 바쁘다는 핑계로 아빠의 몫을 못해 항상 미안하게 생각하는 늙은한 상환, 귀여운 상민과 함께 이 논문의 완성된 기쁨을 나누고 싶습니다.