

논문 2006-43SP-2-7

해마신경망을 이용한 관심 객체 기반의 효율적인 멀티미디어 검색 시스템의 개발

(The Development of Efficient Multimedia Retrieval System of the
Object-Based using the Hippocampal Neural Network)

정 석 훈*, 강 대 성*

(Seok-Hoon Jeong and Dae-Seong Kang)

요 약

본 논문에서는 해마신경망(HCNN:HippoCampal Neural Network)을 이용하여 사용자 친화적인 객체 기반 멀티미디어 검색 시스템을 제안한다. 내용 기반 검색(Content-based Retrieval)에 관한 대부분의 기존의 질의 방법은 입력 영상에 의한 질의 또는 컬러(color), 형태(shape), 질감(texture)등과 같은 low-level의 특징을 사용한다. 본 논문에서 제안하는 방법은 MPEG 기반의 압축 비디오 스트림으로부터 장면 전환 검출을 수행하여 샷을 검출한다. 이 샷 프레임에서 컬러 객체의 자동 추출을 위하여 similar colorization과 ACE(Adaptive Circular filter and Edge) 알고리즘을 사용한다. 그리고 이렇게 추출된 특징을 해마 신경망을 통하여 학습한 후 멀티미디어 검색 시스템을 구성한다. 제안 하는 해마 신경망은 호감도 조절에 의해서 입력되는 영상 패턴의 특징들을 흥분학습과 억제학습을 이용하여 불필요한 특징은 억제시키고 중요한 특징은 흥분학습을 통해 장기기억 시켜서 적응성 있는 실시간 검색 시스템을 구현한다.

Abstract

In this paper, We propose a user friendly object-based multimedia retrieval system using the HCNH(HippoCampus Neural Network). Most existing approaches to content-based retrieval rely on query by example or user based low-level features such as color, shape, texture. In this paper we perform a scene change detection and key frame extraction for the compressed video stream that is video compression standard such as MPEG. We propose a method for automatic color object extraction and ACE(Adaptive Circular filter and Edge) of content-based multimedia retrieval system. And we compose multimedia retrieval system after learned by the HCNH such extracted features. Proposed HCNH makes an adaptive real-time content-based multimedia retrieval system using excitatory learning method that forwards important features to long-term memories and inhibitory learning method that forwards unimportant features to short-term memories controlled by impression.

Keywords : MPEG, Neural Network, Content-based Retrieval, Object Extraction

I. 서 론

최근 디지털 기술의 발전은 영상과 음향 데이터를 통신 및 컴퓨터와 결합하여 새로운 멀티미디어로 발전하고 있다. 이들 중에서 비디오와 관련된 연구로는 비디오 색인 및 검색에 관한 연구가 많이 진행되고 있다^{[1][2]}. 기존에는 파일 이름, 주석, 키워드 등을 특징으로

하는 텍스트 기반의 검색이 사용자들의 접근 용이성을 이유로 사용되어 왔다. 하지만 이런 텍스트 기반의 영상 검색은 거대해져가는 데이터베이스 각각의 멀티미디어에 대한 수작업의 불편함과 엄청난 작업시간, 동일한 멀티미디어에 대한 각기 다른 해석 등, 텍스트를 대표하는 적합한 키워드 선택의 어려움 등을 이유로 많은 문제점을 가지고 있다^[3]. 이에 대한 대안으로 현재 활발하게 연구되고 있는 내용 기반 검색 기법은 영상의 색상(color), 모양(shape), 질감(texture)등의 내용적인 특징이나 속성을 이용하여 영상을 표현하고 이것을 기반

* 정회원, 동아대학교 전자공학과
(Dept. of Electronics, Dong-A University)
접수일자: 2005년8월15일, 수정완료일: 2006년2월27일

으로 검색을 수행한다^{[4][5]}.

본 논문에서는 이런 내용기반 검색을 위하여 비디오 스트림을 분석하여 샷을 찾아낸다. 이 샷 프레임들 similar colorization 시키고 ACE(Adaptive Circular filter and Edge) 알고리즘을 이용하여 객체를 추출한다. 그리고 IV장에서 제안하는 해마신경망(HCNN : HippoCampus Neural Network) 학습 알고리즘을 이용하여 학습한 후 실시간으로 검색 가능한 멀티미디어 검색 시스템을 제안한다.

본 논문의 구성은 다음과 같다. I 장 서론에 이어, II 장에서는 비디오 스트림의 장면 전환 검출 및 대표 프레임 추출 알고리즘에 대해 기술하고 III장에서는 similar colorization과 ACE 객체 추출 알고리즘에 대하여 기술하고 IV장에서는 해마신경망을 이용한 학습과 멀티미디어 검색 시스템의 구현에 대하여 기술한다. V 장에서는 제안한 방식을 적용한 실험 및 결과를 보이고, 끝으로 VI장에서는 결론을 맺는다.

II. 비디오 스트림에서 장면 전환 검출

1. MPEG의 배경이론과 DC영상의 추출

MPEG 비디오 시퀀스에서는 I(Intra), P(Predictive), B(Bidirectionally predictive) picture가 있다^[6]. I picture는 inter frame 예측을 사용하지 않고 해당화면 정보만을 부호화하는 화면으로서 모든 MB(MacroBlock) type은 intra frame이며, 채널 전환 시의 원영상 복구와 오류의 전파를 막기 위해 GOP(Group of Picture)내에 최소 한 장의 I picture를 필요로 한다. P picture는 I 혹은 P picture로부터의 순방향 움직임 보상 예측 수행으로 생기는 화면으로서 MB type은 intra frame, forward inter frame, backward inter frame, interpolative inter frame 예측부호화가 수행된다. B picture가 삽입됨으로써 화면 처리 순서가 원 화면 순서와 달라져서 부호기에서는 B picture를 건너 뛰어 다음의 I, P picture를 우선 부호화하고, 그 후 사이에 있는 B picture를 부호화한다. MPEG 비디오의 시퀀스 층은 다수개의 GOP로 되어있다. 각 GOP는 다수개의 picture로, picture층은 다수개의 슬라이스로, 슬라이스 층은 다수개의 MB로, MB는 휘도 신호 블록과 색차 신호 블록으로 구성되어 있다. 보통 한 GOP당 15개 이하의 picture들로 구성된다. 이때 인트라 부호화인 I picture에서 각 슬라이스를 추출하고 슬라이스의 MB에서 각각 8 x 8 DCT영역의 DC성분을 추출한다. 이때 한 I picture당 1개의 DC영상

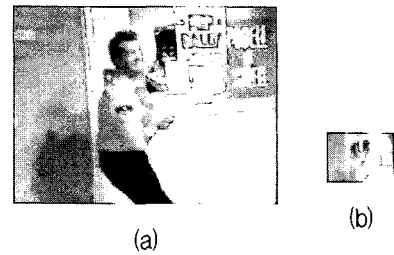


그림 1. MPEG 으로부터 추출된 DC 영상
(a) 원영상(I picture), (b) DC 영상

Fig. 1. DC image from MPEG video stream.
(a) original I picture, (b) DC image

이 생성된다. 본 논문에서는 I picture에 대해서 DCT DC계수를 구하고 이를 정규화 시킨 후 DC영상을 구성하였다. 그림 1은 MPEG 비디오 스트림으로부터 구성된 I picture에 대한 DC영상을 나타낸 것이다.

2. 장면 전환 검출 알고리즘

다음의 3가지 특징들을 이용하여 DC영상에서의 장면 전환 검출을 구하였다^{[7][8]}. 첫 번째 특징은 이전 영상과의 pixel간의 차분 값으로서 다른 영상의 유사도를 측정하는데 가장 기본이 되며, 전체적인 휘도 변화를 잘 나타낸다.

$$DifImg_i = \frac{\sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |I_{i-1}(x, y) - I_i(x, y)|}{W \times H} \quad (1)$$

여기서 W는 DC영상의 가로 픽셀의 수이고, H는 세로픽셀의 수이다.

두 번째 특징은 DC영상의 히스토그램에 대한 chi-square 값이다. 이전 영상에 대한 히스토그램의 변화를 나타낸다.

$$X_i^2 = \sum_{k=0}^{n-1} \frac{(H_{i-1}(k) - H_i(k))^2}{H_i(k)} \quad (2)$$

여기서 $H_i(k)$ 는 i 번째 DC영상 히스토그램의 k 번째 bin의 값을 나타낸다. n은 8bit 영상일 경우 256의 값을 가진다.

세 번째 특징은 이전 DC영상과 현재 DC영상과의 히스토그램 분산의 차분 값이다. 이는 히스토그램의 전체적인 분포에 대한 변화를 나타낸다.

$$DifD_i^2 = \frac{(D_{i-1} - D_i)^2}{D_i} \quad (3)$$

여기서 D_i 는 i 번째 DC영상의 히스토그램 분산 값이다. 이상과 같이 구해진 DC영상의 파라미터들로부터 다음의 단계로 장면 전환 프레임 추출한다.

STEP 1. 각 파라미터들의 전체 프레임에 대한 평균을 구한다.

$$\overline{DifImg}, \overline{X}, \overline{DifD}$$

STEP 2. DC영상의 각 특징 값이 아래의 조건식을 만족하면 장면 전환 프레임으로 검출한다.

condition 1. $DifImg_i > \overline{DifImg}$

condition 2. $X_i > \overline{X}$

condition 3. $DifD_i > \overline{DifD}$



그림 2. MPEG 으로부터 추출된 장면 전환 프레임
Fig. 2. The extracted shot frame from the MPEG.

III. 객체 추출 알고리즘

MPEG은 높은 압축률을 위해 YUV의 좌표계를 가지고 있다. 본 연구에서는 컴퓨터 환경 상에서 주로 표현되는 RGB의 좌표계로 변환한 후에 제안하는 similar colorization 알고리즘을 통하여 영상을 단순화 시키고 ACE 알고리즘을 이용하여 객체를 추출한다.

1. Similar colorization

제안하는 알고리즘은 크게 3부분으로 나뉜다. 우선은 RGB 색채 모델을 HSV 색채 모델로 변환시키고 그 다음으로 HSV 모델을 컬러 양자화 시킨다. 마지막으로 양자화 된 영상의 에지 강화를 위해서 미디언 필터링을 수행한다. 이는 이후의 객체 추출을 위한 알고리즘을 사용할 때 좋은 특성을 제공한다.

가. 색채 모델 변환

RGB 컬러 공간은 디지털 이미지를 위한 가장 일반적

인 컬러 형태이다. 하지만 RGB 컬러는 지각적으로 균일하지 않다는 결점을 가지고 있고 이 때문에 RGB 공간에서의 영상의 양자화는 지각적으로 불필요한 bin들을 많이 가지고 있고, RGB 영역에서 정의된 거리 함수들은 Red, Green, Blue 상호관계를 고려할 때 양자화에는 적합하지 못하다. 그래서 제안한 시스템에서의 영역 변환 방식은 HSV 영역이다. HSV 영역은 인간의 시각 능력과 유사한 특성을 갖는 것으로 알려져 있는데, 각각 색상, 명도, 채도로 대표되는 3가지 요소로 구성되어 있어서 RGB 영역에서와 같이 채널 간 상호관계가 크게 영향을 미치지 않기 때문에 각각 분리된 채널에서의 컬러 영상 처리에 용이하다. 그리고 변환할 때 비선형 적이긴 하지만 비교적 쉽게 전환이 가능하다는 장점이 있다^[9].

HSV 영역에서의 컬러 양자화는 유사하지 않은 컬러를 같은 bin으로 맵핑하는 것을 회피하고 대체로 인간 인지학적 관점으로 영상을 구분할 때 효율이 떨어지는 밝기와 채도영상의 명암도와 채도의 허용 오차에 대한 관용을 비교적 크게 증가 시켰다. 컬러 양자화는 John R. Smith와 Shih-Fu Chang이 제안한 방법을 사용하여 hue는 20도씩 18단계로 되었으며 saturation과 value는 그림 3과 같이 3단계로 양자화 하였다^[10].



(a) value = 0.33 (b) value = 0.67 (c) value = 1.0

그림 3. HSV영역에서 양자화 된 영상

Fig. 3. Quantized Image from HSV color space.

나. 컬러 미디언 필터링

양자화 된 영상에는 여전히 잡음 등이 남아있고 이를 해결하기 위해서 HSV 각 채널에 대해서 5x5 컬러 미디언 필터를 사용한다. 컬러 미디언 필터링의 큰 장점은 강한 에지를 보존하고 기존의 에지들을 좀 더 상세하게 보존할 수 있어서 경계구분에 효과가 있다. 인간의 시각은 이러한 고주파 정보에 민감하다. 이러한 상세한 에지의 보존과 강화 가능성은 이후의 객체 추출을 위한 알고리즘 적용 시 탁월한 성능의 개선을 제공한다.

2. ACE (Adaptive Circular filter and Edge) 객체 추출 알고리즘

사용하는 객체 추출 방법은 크게 2가지 부분으로 나뉜

다. 첫 번째 단계는 세 가지 다른 종류의 적응적인 원형 필터를 사용하여 영상의 질감 정보에 따라 영상의 일차적인 객체를 추출하는 것이고, 두 번째(반복) 단계는 영상의 분할된 작은 부분들을 경역 병합과 영역 레이블링 알고리즘을 이용하여 인근의 유사한 영역들과 병합하는 단계이다^[21].

가. Edge detection and Circular Filter

사람과 가장 유사한 지각을 보여주는 H(Hue)의 값을 이용하여 객체 추출 알고리즘을 적용한다. 그림 4는 원형 필터링을 적용하는 예이다^[21].

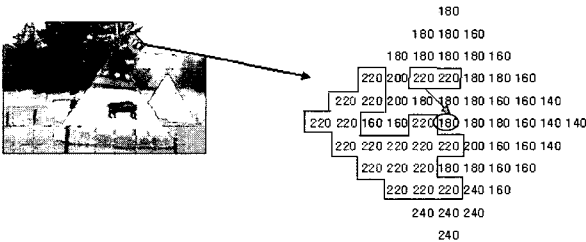


그림 4. 원형 필터의 적용 예
Fig. 4. Applied example of a circular filter.

위의 그림과 같이 원형 필터에 속하는 픽셀들의 값을 카운트하여 가장 많은 수를 차지하는 픽셀의 값을 중간 픽셀 값으로 바꾸어주고 전 영역에 걸쳐 이 과정을 반복한다.

우선 similar colorization 처리된 영상에서 H의 값을 이용하여 에지의 분포를 구하기 위하여 Canny edge detector를 사용한다. 일단 에지가 구해지면 이를 기반으로 9개의 같은 크기의 1st영역으로 균등 분할한다. 여기서 1st영역들의 하위블록의 설정을 위하여 각 영역의 에지 픽셀들의 합계를 계산한다. 그 합계가 같은 하위블록에 포함되어있는 전체 픽셀들의 수의 20%를 초과하면, 그 때 1st영역을 다시 같은 크기의 4개의 2nd영역으로 균등 분할한다. 하위블록의 크기가 가장 큰 원형 필터의 크기보다 적을 때까지 이 처리 과정을 반복한다. 보통 3rd 영역이 적당하다. 이렇게 에지의 수에 따라 영역이 나뉘면, 각 원형 필터를 영역의 수에 따라 하위블록에 적용한다. 예를 들어, 3rd영역이 가장 큰 에지 분포를 가진다면 15×15 원형 필터를 적용하고, 2nd영역에는 11×11 원형 필터를 적용한다. 그리고 마지막으로 1st영역에는 7×7 원형 필터를 적용한다. 마지막으로 영역의 평균 S_m^i 의 유사한 컬러 특성을 갖는 작은 부분들 (S)의 인근영역으로의 병합을 위하여 영역 병합 알고리즘을 사용

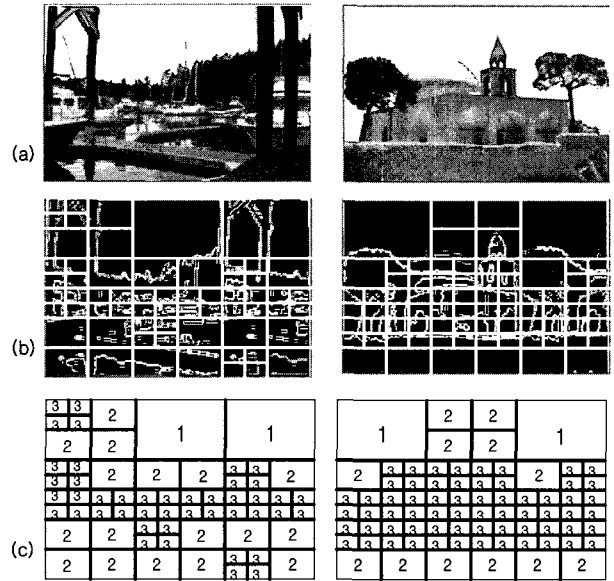


그림 5. 에지에 따라 나뉘지는 하위블록의 종류.
Fig. 5. A kind of subblocks divided according to edges.

한다. 이 알고리즘은 식 (4)에 의해 수행된다^[21].

$$\sum_i^N (|S_H - S_{mH}^i| + |S_S - S_{mS}^i| + |S_V - S_{mV}^i|) \leq T \quad (4)$$

여기서 N은 이웃한 영역들의 숫자이고 T는 영역 병합을 위한 최소 임계치(threshold)이다. 첫 번째 단계의 분할 결과로는 초기에 임의로 각각의 영역으로 나누었기 때문에 각 영역들 상호간의 유사성은 비교 될 수 없었다. 즉, 각 영역들의 내부는 고려되었지만 영역들 간의 유사도는 고려되지 않았다. 그래서 두 번째의 단계는 이를 보완하기 위하여 전역적인 반복 병합 알고리즘을 적용한다.

나. 영역 간 병합과 레이블링 알고리즘

우선 같은 컬러 값을 가지는 각 영역은 특정한 값으로 버퍼에 저장한다. 이러한 이진 영상은 왼쪽 위에서 오른쪽 아래로 가면서 raster scan 방식으로 각각 스캔한다. 이 단계에서 각 영역에 따라 다른 인덱스가 할당된다^[21]. 영역 레이블링 이후, 영역들의 수가 30개 이상이 되면, 우리는 첫 단계에서 병합되지 않았던 인근 영역들로의 병합을 위하여 가장 작은 7×7필터를 이용하여 원형 필터링을 반복하고, 임계치 T로 영역 병합을 수행한다. 임계치 T는 영역의 수가 30개 이하로 감소 할 때까지 계속해서 반복함으로써 증가한다. 더군다나, 본 논문의 실험에서 같은 영역의 크기(픽셀의 수)가 영상의 0.3% 보다 적으면 그것은 또한 가장 큰 이웃한 영역으로 병합되었다. 이렇게 병합되고 인덱스 된 정보를 이용하여 다음 장

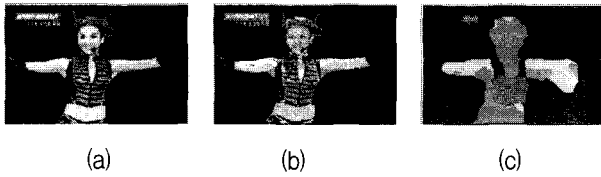


그림 6. (a) 원영상, (b) similar colorization, (c) ACE
 Fig. 6. (a) Original Image, (b) similar colorization, (c) ACE.

에서 설명할 해마 신경망을 이용하여 학습한다.

그림 6의 (a), (b), (c) 는 각각 원 영상을 similar colorization한 영상과 ACE 알고리즘으로 객체를 추출한 영상이다. 그림에서 보는 바와 같이 ACE 알고리즘을 이용한 처리를 통하여 컬러의 bin수가 현저히 줄고 질의하는 객체가 또한 뚜렷한 군더더기 없는 영상으로 처리되었음을 알 수 있다.

IV. 해마신경망을 이용한 검색 시스템 구현

1. 해마신경망

(HCNN : HippoCampus Neural Network)

사람의 뇌 속에서 실제 어떠한 패턴을 인식할 때는 해당 패턴의 특이한 특징에 대해 흥분하는 세포들이 특정 신호를 발생시킨다. 예를 들어, 사과를 보았을 때 빨간색에 흥분하는 세포와 동그란 원형에 흥분하는 세포 등이 각각의 신호를 발생시킨다. 이러한 정보가 해마에서 분석되어 저장된 기억 속에 사과의 맛을 찾아내어 입에 침이 고이도록 만든다. 이때 뇌 속에 기억되는 정보는 해당 특징에 흥분한 세포들의 조합된 결과 하나로서만 기억되는 것이다. 일반적인 신경망에서 사용하는 인식 방법과의 차이점은 "A" 라는 문자와 "B" 라는 문자를 구분하기 위해 문자를 위해 추출한 모든 특징점(좌우 대비, 길이, 교차점 개수 등)을 모두 사용하여 정보를 저장시킨다는 것이다. 그러나 뇌에서는 실제로 해당 패턴 하나에 필요한 흥분 세포들의 정보만 저장된다. 일반적인 신경망 구조를 갖는 체제에서는 필요한 특징들만 저장할 수 없다. 본 논문에서는 해당 특징에 따라 유동적으로 정보를 저장할 수 있는 시스템을 제안할 것이다. 뇌 속의 해마를 통해 분석된 정보는 편도체에서 좋아하는 정보로 판단되면 학습이 쉽게 이루어진다. 반면에, 좋아하지 않는 정보는 휘발성 영역인 단기 기억 장소에 기록되고, 여러 번의 반복을 통하여 잊혀지는 시기가 늦추어지는 장기 기억 장소에 기록된다. 호감도가 높은 정보에 대해서는 장기 기억 장소에 해당하는 영역으로 저장시키고, 호감도가 낮

은 정보에 대해서는 단기 기억장소에 일단 보관하여 여러 입력 패턴에서 지속적으로 패턴에 대해서 호감도가 높아지면 장기 기억 장소에 저장시킬 수 있을 것이다^{[12]-[14]}. 우리는 이러한 호감도 조절을 통해서 단기와 장기 기억에 따른 정보 저장을 제안할 것이다.

2. HCNN 학습 알고리즘

해마의 기능 중 가장 중요하고 강력한 것은 단기기억에서 장기기억으로의 확장에 있다. 해마의 역할은 방금 저장된 정보를 기억하고 필요한 정보와 그렇지 못한 정보를 분류하는 것이다. 필요한 정보의 경우, 해당 뉴런에 저장해두고 동일한 패턴이 입력되면 인식하게 되는 것이다. 단기 기억과 장기 기억으로 나누는 방법에 있어서 호감도 측정은 통계적 패턴 인식에 사용되는 방법을 사용해서 통계적으로 많이 나타나는 값들에 대해서 호감도를 높게 부여하는 것이다. 이러한 특징을 해마 신경망에 도입한다면, 입력되는 중간에 입력받는 값들의 가중치를 사용하여 단기 기억을 할 것인지 장기 기억을 할 건인지를 결정할 수 있다. 이러한 결정 값을 해마기억 장소 분류 기준 값으로 사용할 수 있다. 전체 데이터 개수 대비 해당 뉴런의 누적 개수가 문턱치(threshold)를 만족하면 장기 기억 장소로 저장시키도록 한다.

일반 신경망은 동적으로 확장되지 못하는 구조적인 한계 때문에 에러율이 고정될 수밖에 없다. 실험을 통해서도 $M * M$ 크기의 구조보다 $(M+1) * (M+1)$ 의 구조가 인식률이 높아지는 것은 고정된 M 의 크기에 대해서 인식률이 영향을 받는다는 것이다. 해마 신경망은 인식할 시스템에서 사용할 메모리 크기가 한정되는 경우 높은 인식률을 위하여 입력되는 특징점의 누적 개수가 많은 데이터를 장기 기억장소에 기억시킨다. 단기기억에서 장기기억으로의 확장을 위하여 본 논문에서는 중요한 특징에 대한 가중치 부과를 통하여 인식률을 높이고 적응적으로 확장 가능한 신경망 학습 알고리즘을 제안한다. 해마 지역은 뉴런의 가중치의 모듈레이터 역할을 한다. 변조 가중치 v_i 를 통해 식 (5)에 의해서 변조된다^[15].

$$v_i = \frac{2}{1 + e^{(-\sum_{k \in C_k} m_k(d) c_k(d))}} - 1 \quad (5)$$

해마 뉴런의 변조요소는 입력되는 각 패턴의 동일성을 나타내는 C_k 와 주어진 threshold 값에 의해 식 (6), 식 (7)과 같이 갱신된다. 갱신된 변조요소는 식 (6)과 같은 흥분 학습과 식 (7)의 억제 학습을 통하여 변조 가중치를

변화 시켜서 상대적으로 중요한 패턴은 장기기억으로 학습시키고 불필요한 패턴의 가중치는 비활성 시킨다^[16].

$$m_i(t+1) = m_i(t) + \alpha \left(\frac{c_{ki}(t) - s}{s} \right) \quad (6)$$

$$m_i(t+1) = m_i(t) + \alpha \left(\frac{c_{ki}(t) - s}{1 - s} \right) \quad (7)$$

3. 멀티미디어 검색 시스템의 구현

질의 영상에 대한 검색의 결과로써, 그에 상응하는 동영상의 호감도, 즉 가중치가 높은 순서대로 플레이가 가능하게 한다. 이를 위해서는 II장에서 소개한 장면 전환 프레임의 검출 시에, 추출되는 샷에 대한 시간 정보가 필요하게 된다. 본 시스템은 크게 3가지로 구분된다. 우선은 MPEG 비디오 스트림에서 샷을 검출하는 시스템, 두 번째로 이렇게 검출된 샷을 이용하여 학습을 하는 시스템 그리고 마지막으로 질의 영상을 이용하여 원하는 멀티미디어를 검색하는 시스템으로 구성된다. 그러므로 샷 검출 시에 그때의 멀티미디어가 디스 플레이된 시간의 정보를 알아내어 저장함으로써 구현이 가능해진다.

V. 실험 및 결과

1. 실험 환경

제안한 멀티미디어 검색 시스템을 실험하기 위하여 Intel Pentium 4 2.00 GHz PC의 Microsoft Windows XP 환경에서 실험하였다. 그리고 Visual C++ 6.0 MFC를 이용하여 검색 시스템을 개발하였다. 실험 데이터는 검색의 편의를 위하여 오디오 성분이 제거된, 192×144 크기의 MPEG 비디오 스트림을 사용 하였다. 실험은 10개의 뮤직비디오 영상, 5개의 뉴스 영상, 3개의 드라마 영상 그리고 5개의 풍경 영상으로 구성 되어있으며, 각 영상들에서 추출한 대표 프레임 이미지의 수는 총 428개이고, 이를 이용하여 데이터베이스를 구축하였다. 다음의 표 1

표 1. 실험에 사용한 멀티미디어에 대한 정보
Table 1. Information about the Multimedia using the experiment.

종류 (개)	총 시간 (분:초)	추출된 프레임 수 (개)
뮤직비디오 (10)	25 : 33	237
뉴스 (5)	14 : 46	81
풍경 (5)	21 : 25	67
드라마 (3)	13 : 39	43

은 본 시스템에서 사용한 동영상에 대한 정보이다.

2. 실험 결과

그림 7은 대표 프레임과 그 프레임에 대한 시간 정보를 추출하는 프로그램이다.

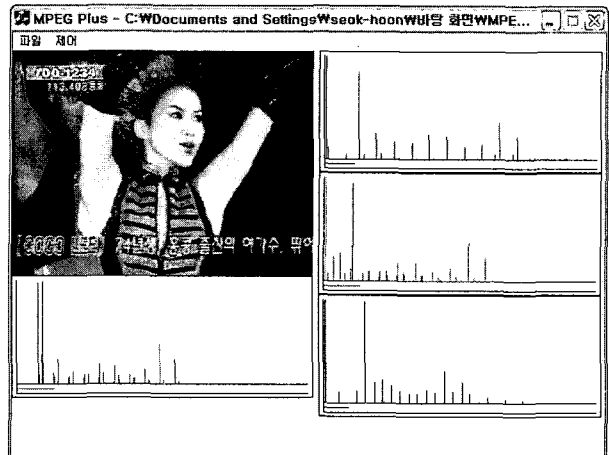


그림 7. MPEG으로부터 대표 DC 프레임 추출.

Fig. 7. Extracted the representative DC frame from MPEG/

이렇게 추출된 대표 프레임을 이용하여 앞에서 소개한 방법으로 객체를 추출한다. 본 논문에서는 최근 많이 사용되고 있는 EM(Expectation-Maximization)^[17] 및 VQ(Vector Quantization)^[18]의 클러스터링 알고리즘과 소요시간을 비교 분석하였다. EM의 경우 각 영상을 구성하는 객체 수에 알맞게 클러스터의 개수를 자동적으로 결정짓지 못하기 때문에 각 영상마다 2~7개의 클러스터로 분류하는데 걸리는 시간을 계산하였다. 다음의 표 2는 객체를 클러스터링 하는데 걸리는 시간을 비교한 것이다.

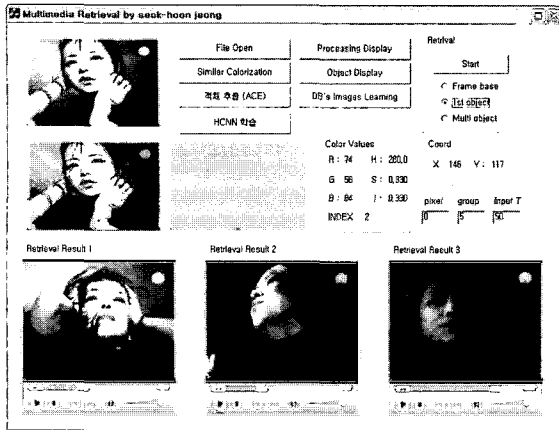
이렇게 추출된 특징 비트 스트림들은 IV절에서 제안한 해마신경망(HCNN)으로 학습된다. 그림 8은 사용자가 질의한 영상을 이용하여 원하는 멀티미디어를 검색하는 시스템이다.

본 연구에서는 10개의 뮤직비디오와 5개의 뉴스, 3개의 드라마 그리고 5개의 풍경 동영상을 이용하여 각 동영상에서 추출된 428개의 대표 프레임으로 검색을 수행하였다. 실험 결과의 검색 성능을 비교하기 위하여 다음의

표 2. 클러스터링의 소요시간 비교
Table 2. A period comparison of clustering algorithms.

이미지 수 (개)	EM (초)	VQ (초)	ACE (초)
100	369	263	153
200	612	429	302
428	1336	782	643

그림 8. 선택 객체에 기반한 검색 결과
Fig. 8. The retrieval result based on the selected object.



식 (8), 식(9)와 같이 Recall과 Precision을 사용 하였다.

$$Recall = \frac{\text{정확하게 검색된 영상수}}{\text{관련된 모든 영상수}} \quad (8)$$

$$Precision = \frac{\text{정확하게 검색된 영상수}}{\text{검색된 모든 영상수}} \quad (9)$$

본 논문에서 구현한 검색 시스템의 성능을 측정하기 위해서 실험에 사용된 동영상들에서 추출된 대표 프레임 들을 이용하여 Jing huang에 의해 제안된 color correlogram과 Jinsham Tang, Scott Acton이 제안한 multi-example retrieval 알고리즘과 비교 실험을 하였다. color correlogram은 색상의 공간분포를 이용하는 방법으로 좋은 검색 결과를 보여주고 있고, 본 데이터베이스 를 이용하여 실험한 multi-example retrieval 시스템은 영상의 컬러 히스토그램과 질감 히스토그램을 특징으로

표 3. 선택된 객체에 기반한 검색 성능 분석 결과
Table 3. Recall and precision comparison based on selected object.

분 류	color correlogram		multi-example		HCNN	
	Recall	Precision	Recall	Precision	Recall	Precision
뮤직비디오	78.2	70.8	80.3	72.4	87.4	80.2
뉴스	85.1	76.2	83.2	74.2	91.8	85.4
드라마	82.2	75.6	84.5	75.1	88.4	81.9
풍경	80.1	72.7	81.3	69.6	89.7	83.4
평 균	81.83	74.20	82.67	73.90	89.20	82.50

사용한 알고리즘으로 기존의 컬러 중심의 내용 기반 검색보다 높은 검색률을 가지고 있다^{[19][20]}. 다음의 표 3은 사용자가 선택한 임의의 관심 객체에 기반한 검색 시스템의 성능을 분석한 결과이다.

VI. 결 론

본 논문에서는 객체에 기반한 내용 기반 멀티미디어 검색 시스템을 구현하였다. 제안한 시스템은 우선 MPEG 동영상 스트림에서 대표 프레임 샷을 추출해 낸다. 멀티미디어 검색에서는 정확한 샷을 검출할 필요가 없으므로 확연한 차이가 나는 샷만을 대표 프레임으로 선택한다. 그리고 이렇게 추출된 대표 프레임을 similar colorization을 통하여 단순화시키고, ACE 알고리즘으로 보다 유연한 객체를 추출한다. 이렇게 추출된 객체를 이용하여 앞에서 제안한 해마신경망을 이용하여 학습하여 효과적이고 속도가 향상된 검색 시스템을 제안 하였다. 앞으로의 계획은 더 많은 데이터베이스를 보유 하여 보다 광범위한 범위에서 정확도 높은 검색 성능을 가지는 시스템을 구현할 것이며, 객체뿐만 아니라 컬러 상관도와 공간좌표 등에 기반한 검색 알고리즘을 보충하여 동적인 다중 멀티미디어 검색 시스템을 구현할 것이다.

참 고 문 헌

- [1] Wei Xiong and Chung-Mong Lee, "Efficient Scene Change Detection and Camera Motion Annotation for Video Classification", Computer Vision and Image Understanding Vol.71, No. 2/2, August, pp.166-181, 1998.
- [2] Hong Heater Yu, "A hierarchical Multiresolution Video shot Transition Detection Scheme", Computer Vision and Image Understanding Vol.75, No. 1/2, July/August, pp. 196-213, 1999.
- [3] A. Vailaya, A. K. Jain and H. J. Zhang, "On Image Classification: City Images vs. Landscapes", Pattern Recognition, Vol.31, pp. 1921-1936, 1998.
- [4] B. Manjunath and W. Ma, "Texture features for browsing and retrieval of image data," IEEE Trans. Pattern Anal. Machine Intel, vol 18, pp. 837-842, Aug. 1996.
- [5] R. Mehrotra and J. Gary, "Similar-shape retrieval in shape data management," IEEE Computer, vol. 28, pp. 57-62. Sept. 1995.

- [6] 후지와라 히로시 "그림으로 보는 최신 MPEG", 교보 문고, 2001.
- [7] John S. Boreczky and Lawrence A. Rowe, "Comparison of Video Shot Boundary Detection Techniques", Storage and Retrieval Image and Video Database IV, Proc. of IS&T/SPIE 1996 Symp. on Elec. Imaging: Science and Technology, February 1996.
- [8] Arun Hampapur, Ramesh Jain and Terry Weymouth, "Digital video Segmentation", Proc. Second Annual ACM Multimedia Conference, October. 1994.
- [9] R. C. Gonzalez, R.E.Woods, *Digital image processing*, Prentice-Hall, 2001.
- [10] John R. smith and Shih-Fu Chang, "Tools and Techniques for Color Image Retrieval", IS&T/SPIE proceedings vol. 2670, Storage & Retrieval for Image and Video Database, 1995.
- [11] Kishan Mehrotra, Chilukuri K. Mohan and Sanjay Ranka, *Elements of Artificial Neural Networks*, The MIT press, 1997.
- [12] 이케가야 유지, 이토이 시게사토, 해마, 은행나무, (2003).
- [13] Dayan, P. and Abbott, L.F., *Theoretical Neuroscience*, MIT press, 2001.
- [14] Ventriglia, F. and Maio, V.D., "Synaptic fusion pore structure and AMPA receptor activation according to Brownian simulation of glutamate diffusion, *Biological Cybernetics*", Vol. 88, No, 3, 2003.
- [15] D.G. Amaral and M. P. Witter. "The three-dimensional organization of the hippocampal formation: A review of anatomical data," *Neuroscience*, vol. 31, pp.571-591, 1989.
- [16] R. Miller, "Cortico-Hippocampal interplay and the representation of contexts in the brain. Springer Verlag. 1991. technique for OFDM and MC-CDMA in a multipath fading channels," in Proc. of IEEE Conf. on Acoustics, Speech and Signal Processing, pp. 2529-2532, Munich, Germany, May 1997.
- [17] Serge Belongie, Chad Carson, Hayit Greenspan, and Jitendra Malik, "Color and Texture-Based Image Segmentation Using EM and Its Application to Content-based Image retrieval," Sixth International Conference on Computer Vision, pp. 675-682, January. 1998.
- [18] Abhijit. S. Pandey, *Pattern Recognition With Neural Networks in C++*, IEEE Press, 1995
- [19] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, and R. Zabih, "Image Indexing Using Color Correlograms," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 762-768, 1997.
- [20] Jinshan Tang and Scott Acton, "An Image Retrieval Algorithm using Multiple Query Images," IEEE Proc. Signal Processing and Its Applications, vol. 1, pp. 193-196, 2003.
- [21] B. Ko, H. Byun, "FRIP: a region-based image retrieval tool using automatic image segmentation and stepwise Boolean AND matching" IEEE Trans on Multimedia, Vol.07, No.01, pp.0105~0113, 2005.02.

 저 자 소 개



정 석 훈(정회원)
 2004년 동아대학교 전자공학과
 학사 졸업.
 2006년 동아대학교 전자공학과
 석사 졸업.
 <주관심분야 : 내용 기반 멀티미
 디어 검색, MPEG, 신경망>



강 대 성(정회원)
 1984년~1989년 국방과학연구소
 연구원
 1991년 Texas A&M 대학교
 전자공학과 석사 졸업.
 1994년 Texas A&M 대학교
 전자공학과 박사 졸업.
 1994년~1995년 ETRI 선임연구원
 2006년 현재 동아대학교 전자컴퓨터공학부
 부교수.
 <주관심분야 : 영상처리, 비디오 색인 및 검색, 영
 상 압축>