대화형 유전자 알고리즘을 이용한 감성기반 비디오 장면 검색

(Emotion-based Video Scene Retrieval using Interactive Genetic Algorithm)

유 헌 우 † 조 성 배 ††
(Hun-Woo Yoo) (Sung-Bae Cho)

요 약 본 논문에서는 감성에 기반한 장면단위 비디오 검색방법을 제안한다. 먼저 특정 줄거리를 담은 장면 비디오 클립에서 급진적/점진적 샷 경계 검출 후, "평균 색상 히스토그램", "평균 밝기", "평균 에지 히스토그램", "평균 샷 시간", "점진적 샷 변화율"의 5가지 특징을 추출하고, 이 특징과 사람이 막연하게 가지고 있는 감성공간과의 매핑을 대화형 유전자 알고리즘(IGA, Interactive Genetic Algorithm)을 통하여 실현한다.

제안된 검색 알고리즘은 초기 모집단 비디오들에 대해 찾고자 하는 감성을 내포하고 있는 비디오를 선택하면 선택된 비디오들에서 추출된 특징 벡터를 염색체로 간주하고 이에 대해 교차연산(crossover)을 적용한다. 다음에 새롭게 생성된 염색체들과 특징벡터로 색인된 데이타베이스 비디오들간에 유사도 함수에 의해 가장 유사한 비디오들을 검색하여 다음 세대의 집단으로 제시한다. 이와 같은 과정을 여러 세대에 걸쳐서 실행하여 사용자가 가지고 있는 감성을 내포하는 비디오 집단들을 얻게 된다. 제안된 방법의 효과성을 보이기 위해, 300개의 광고 비디오 클립들에 대해 "action", "excitement", "suspense", "quietness", "relaxation", "happiness" 의 감성을 가진 비디오를 검색한 결과 평균 70%의 만족도를 얻을 수 있었다.

키워드 : 감성기반 검색, 비디오 장면 검색, 대화형 유전자 알고리즘

Abstract An emotion-based video scene retrieval algorithm is proposed in this paper. First, abrupt/gradual shot boundaries are detected in the video clip representing a specific story. Then, five video features such as "average color histogram", "average brightness", "average edge histogram", "average shot duration", and "gradual change rate" are extracted from each of the videos and mapping between these features and the emotional space that user has in mind is achieved by an interactive genetic algorithm.

Once the proposed algorithm has selected videos that contain the corresponding emotion from initial population of videos, feature vectors from the selected videos are regarded as chromosomes and a genetic crossover is applied over them. Next, new chromosomes after crossover and feature vectors in the database videos are compared based on the similarity function to obtain the most similar videos as solutions of the next generation. By iterating above procedures, new population of videos that user has in mind are retrieved. In order to show the validity of the proposed method, six example categories such as "*action*", "*excitement*", "*suspense*", "*quietness*", "*relaxation*", "*happiness*" are used as emotions for experiments. Over 300 commercial videos, retrieval results show 70% effectiveness in average.

Key words : Emotion-based Retrieval, Video Scene Retrieval, Interactive Genetic Algorithm (IGA)

1. 서 론

대용량 저장 시스템의 일반화, 네트워크의 보편화 등 컴퓨터 관련 기술의 진보로 영상과 비디오 데이타를 전송, 처리하고 검색하는 관리 시스템이 필요하게 되었다. 특히 스캐너나 디지털 카메라, VCR 등의 화상입력 장치의 가격 하락으로 일반 대중들도 이러한 멀티미디어

· 본 논문은 2002년도 학술진흥재단의 연구지원에 의해 수행되었음 (KRF-2002-005-H20002)
† 비 회 원 : 연세대학교 인지과학연구소 연구교수
paulyhw@yonsei.ac.kr
†† 종신회원 : 연세대학교 컴퓨터산업공학부 교수
sbcho@yonsei.ac.kr
논문접수 : 2003년 9월 26일
심사완료 : 2004년 9월 13일

데이타를 손쉽게 생산하고 저장하며 웹상에서 동호회를 조직하여 이러한 데이타를 서로 주고 받으며 자신의 개성을 발휘하는 시대에 살게 되었다. 앞으로도 이러한 추세는 계속될 것이며 새로운 시대를 맞이하여 광범위하게 생성되는 멀티미디어 데이타를 사용자의 요구에 맞게 저장, 검색하는 관리 시스템의 개발이 절실히 요구된다. 이러한 시스템은 전자도서관, 홈쇼핑, VOD(Video On Demand) 서비스, 원격진료 등을 실현하는데 필수적이며 교육분야에서도 효과적으로 사용될 수 있다.

이러한 요구를 충족시키기 위해 개발된 초기의 검색 시스템은 대부분 키워드 방식(query by keyword)으로 이루어졌다. 그러나 색인에 대한 부담감 등 여러 가지 단점이 제기 되면서, 최근에는 컴퓨터에 의한 자동 색인 방식을 이용한 내용기반 검색 방식(query by content)이 주목을 받고 있다. 이 방식은 영상에서 색상, 질감, 모양정보, 혹은 객체내의 위치관계 등을 가지고 사용자의 요구에 맞는 데이타를 검색하도록 되어있다[1-8]. 주로 영상처리나 패턴인식, 영상분리, 객체분리 등의 컴퓨터 비전에서 필요한 알고리즘을 사용하여 실현한다. 비디오의 경우는 시간정보를 포함하기 때문에 먼저 샷의 경계를 검출하여 키 프레임을 생성하고, 키 프레임에서 시-공간적인 특징을 추출한 후, 장면단위의 군집화 알고리즘을 적용하여 계층적으로 데이타를 구성 축약(abstract)하여 검색에 이용하는 기술 등이 기본적으로 수반된다[9,10].

그러나 초기의 기대와는 달리 색상, 질감, 모양, 객체의 위치관계 등은 인간이 가지고 있는 풍부한 의미론적 정보(semantic information)를 표현하기에는 아직 미흡하기 때문에 여러 가지 한계에 직면하고 있다. 예를 들어, 사용자가 푸른 하늘을 보여주는 영상과 비디오를 검색하고자 원한다면 색상정보 등을 가지고 검색할 때 하늘 뿐만 아니라 유사한 색상을 지닌 바다도 검색될 것이다. 또한 현재의 컴퓨터비전기술로는 하늘에 떠있는 새와 비행기 조차도 완벽하게 구별하기 어려운 실정이다.

따라서 이러한 단점을 극복하고자 최근에는 영상과 비디오가 가지고 있는 의미정보를 추출하여 의미에 기반한 검색을 수행하는 연구에 초점이 맞추어져 있다[11-15]. 이를 위해서 영상내의 객체를 분리하여 학습하는 방식 [11], 베이즈 이론을 이용하여 영상을 분류하는 방식[12,13], 사용자의 의도를 유사도 피드백을 통하여 실현하는 방식[14,15] 등이 있다.

현대는 감성의 시대라는 말이 있다. 상품 디자인, 영화포스터나 예고편 등을 제작할 때 기술적인 측면뿐만 아니라 사용자의 감성을 고려함으로써 보다 고객에게 다가설 수 있기 때문에, 최근에는 기술보다는 감성적인 측면을 보다 강조하는 추세에 있다. 따라서 감성에 기반한 미디어검색(emotion-based retrieval)은 필수적이며 의미론적 미디어 검색의 한 분야로서 감성을 이용한 검색은 매우 중요한 분야를 차지할 것이다. 아직 컴퓨터 비전분야와 연관 지어서 감성을 기반으로 한 미디어 검색분야는 연구가 활발하지 않다.

감성을 추출하여 검색에 이용하는 방법은 몇몇 연구자들에 의해 시도되었다[18-21]. 조성배는 웨이블릿 계수를 이용하여 우울한 분위기의 영상과 화려한 분위기의 영상을 구별한 후 대화형 유전자 알고리즘(IGA)을 이용하여 여러 번의 피드백을 통해 원하는 영상을 검색하는 방법을 이용하였다[17]. 그러나 이 연구는 실험에서 제한적인 감성을 웨이블릿 계수만으로 구별하였다는 제약이 있다. 비슷한 연구로 Takagi는 인간의 감성에 기반한 검색을 위해 인간의 감성을 표현하는 심리적인 공간(psychology space혹은 factor space)을 설계하고 영상에서 추출한 특징을 표현하는 물리적 공간간의 매핑 관계를 도출한 후 대화형 유전자 알고리즘(IGA)을 통하여 검색을 수행하였다[18].

엄진섭은 Soen[16]의 심리학적인 실험결과로부터 칼라와 그래이 패턴의 물리적 속성과 감성과의 관계를 인공 신경망과 퍼지 이론을 이용하여 감성평가 하는 모델을 제안하였다[19]. 이 방법은 영상에서 칼라와 명도, 텍스쳐 정보를 추출한 후 모델에 입력하면 모델은 13개의 개별 감성에 대해서 강도를 출력하도록 되어있다. 그러나 이 방법은 5개의 영상만을 가지고 테스트함으로써 일반화의 문제점이 있을 수 있고, 영상이 모델에 입력되면 출력으로 개별 감성정보가 추출되므로 반대의 경우인 영상검색에 바로 이용되지는 않았다(영상검색은 입력으로 텍스트로 표현된 감성정보, 예를 들어 "heavy한 영상을 검색해라", 혹은 감성을 내포하는 영상(query by example) 등이 사용되고 출력으로는 입력과 유사한 정보를 가지고 있는 영상들을 얻는 것이다).

Colombo 등은 Itten[22]의 칼라이론을 이용해 예술영상을 감성에 기반하여 검색하였다[20]. 이 방법은 먼저 영상을 유사한 영역으로 구분하고 각 영역에서 색상(color), 더운 정도(warmth), 색조(hue), 밝기(luminance), 채도(saturation), 위치(position), 크기(size)정보를 추출하고, 다른 영역과의 대비(contrast)와 조화(harmony) 등의 관계정보를 Itten의 모델에 근거해서 감성정보로 매핑 시켰다. 그러나 이 방법은 대상을 회화영상에만 국한시켰다는 단점이 있다. 저자는 또한 감성기반 비디오검색방법도 제안했는데, 상업용 광고 동영상을 기호학적인 분류(semiotic category)에 따라 유토피아적인(utopic), 비판적인(critical), 실용적인(practical), 재미있는(playful)의 4가지 방식으로 구별하여 검색하였다[20, 21].

위와 같이 감성관련 연구는 아직 많지 않으며, 특히 비디오의 감성을 표현하고 검색하는 연구는 매우 드물다는 것을 알 수 있다. 따라서 본 연구에서는 장면 단위로 비디오가 가지고 있는 감성정보를 추출하여 감성기반 검색을 실현하는 알고리즘을 제안하고자 한다. 제안된 방법은 검색과정에 사용자의 의도를 반영하게 함으로써 대상이 불명확한 비디오의 장면을 효율적으로 검색할 수 있도록 한다. 기본적으로 장면으로 표현될 수 있는 비디오 클립을 샷 단위로 분할하고 매 샷 마다 키 프레임을 얻은 후, 장면내의 여러 키 프레임에서 평균 색상 히스토그램, 평균밝기, 평균 에지 히스토그램, 평균 샷 시간, 점진적 샷 변화율의 5가지 특징을 추출하여 염색체로 표현하고 데이타베이스에 입력한 후 이를 대화형 유전자 알고리즘을 이용하여 장면단위 감성검색을 실현한다. 대화형 유전자 알고리즘은 가치 평가자로서의 사람과 최적화를 위한 유전자 알고리즘이 조합된 기술로서 컴퓨터 상에서 사용자의 요구 사항이나 기호를 반영할 수 있다. 제안된 방식이 Colombo의 방식과 다른 점은 비디오에서 얻을 수 있는 감성을 특정한 몇 가지로 제안하지 않으며, 대화형 유전자 알고리즘을 이용하여 사용자의 유사도 피드백을 통해 원하는 감성의 비디오를 검색해 낸다는 것이다. 본 논문의 기여도는 아래 3가지로 요약될 수 있다. 첫째, 인간에 감성에 기반한 비디오 검색을 실현하였고, 둘째, 대화형 유전자 알고리즘을 사용하여 목적함수가 명시적으로 정의되지 않은, 즉 사용자가 막연하게 가지고 있는 감성을 내포하거나 대상이 불 명확한 장면 비디오를 검색하는 방법을 제안하였다. 마지막으로, 질의로 사용되는 감성을 특정한 몇 가지로 한정 짓지 않음으로써 일반화의 장점이 있다. 제안된 방법의 전체적인 구성은 그림 1과 같다.

2. 비디오 샷 경계 검출

비디오는 여러 개의 프레임으로 표현된 영상들이 시공간의 압축을 통해서 시퀀스 형태로 저장되어 있다. 따라서 감성에 기반한 비디오 검색을 위해서는 먼저 비디오를 샷 단위로 분할하고 키 프레임을 추출하여 관련 특징을 얻어야 한다. 비디오는 크게 상위 레벨에서 하위 레벨로 비디오 →장면 →샷 →키 프레임의 순으로 이루어져 있다(그림 2와 표 1).

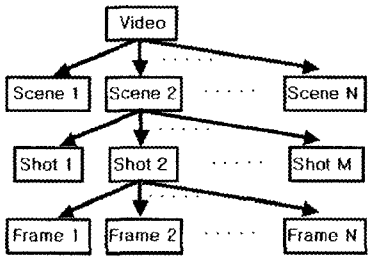


그림 2 비디오의 계층적 구조

표 1 비디오관련 용어

용 어	의 미
장면 (scene)	공통적인 사건이나 시간의 연속성에 의해 결합된 연속된 샷의 집합을 의미한다. 따라서 의미적으로 공통된 특징을 나타내는 줄거리라고 생각할 수 있다.
샷 (video shot)	비디오 제작시 카메라의 기록(record)과 멈춤(stop) 사이의 일련의 연속된 비디오 프레임들의 집합.
키 프레임 (key frame)	여러 개의 프레임으로 구성된 샷에서 가장 중요한 정보를 표현하는 프레임. 정보의 복잡도에 따라 1개의 샷에는 1-2개의 키 프레임을 사용한다. 본 연구에서는 샷의 처음 프레임을 해당 샷의 키 프레임으로 간주한다.

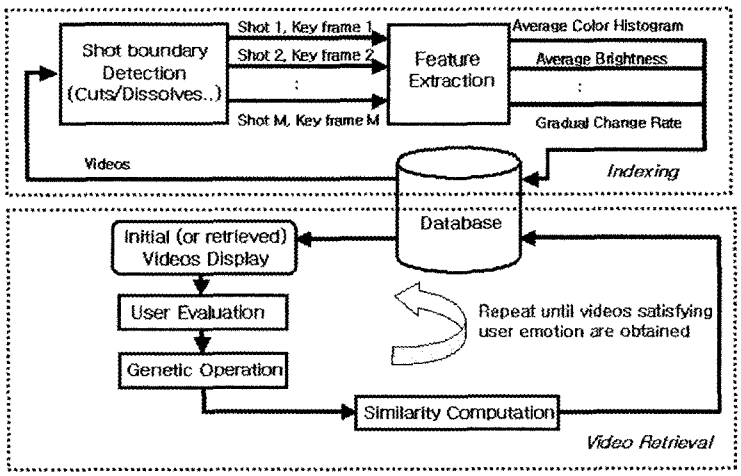


그림 1 제안된 방법

감성에 기반한 장면 단위의 비디오 검색을 위해서는 먼저 비디오의 샷 경계를 검출하여야 한다. 샷 경계에는 컷(cut)으로 표현되는 급진적 변화가 있고, 디졸브(dissolve) 등으로 표현되는 점진적 변화가 있다. 본 연구에서는 이전 연구에서 사용했던 밝기 히스토그램과 에지 개수의 프레임간 상관계수(correlation coefficient)를 통합하여 비디오의 샷을 검출한다[23]. 요약하면, 만약 현재 프레임이 이전 프레임과 충분히 다를 경우(급격한 전환)나 이 전의 샷 경계와 충분히 다를 경우(점진적 전환) 현재 프레임을 샷의 경계프레임으로 판단한다. 충분히 다르다는 것을 판단하는 기준으로는 보통 인접한 각각의 프레임에서 특징(예를 들면 밝기 히스토그램)을 추출하여 특징간의 거리(예를 들면 유클리디안 거리)를 일정 임계값과 비교하여 유사도를 판단하나 이 경우는 다양한 영상에 공통적으로 적용될 수 있는 임계 값 설정이 어려우므로 프레임간의 유사도인 상관계수를 이용하여 모든 비디오에 공통적으로 적용될 수 있는 임계값을 적용한다. 그림 3에서 t_1 은 급진적 전환을 위한 임계 값으로, t_2 는 점진적 전환을 위한 임계 값으로 사용된다. 구체적으로, 제안된 알고리즘은 이전 프레임(p)과 현재 프레임(c) 간의 상관계수 $S_{IFC}(p,c)$를 구한 후, 이를 미리 정한 급진적 전환 검출을 위한 임계값 t_1과 비교하여 작을 경우 급진적 전환 프레임(샷 경계 프레임)으로 간주하고 같거나 큰 경우는 현재 프레임(c)을 이전의 샷 경계 프레임(b)의 상관계수 $S_{IFC}(p,c)$와 비교한다(샷 경계 프레임이 없을 경우 첫 번째 프레임을 디폴트 샷 경계 프레임으로 삼는다). 만약 미리 정한 임계값 t_2보다 작으면 점진적 전환(샷 경계프레임)으로 간주하고 같거나 클 경우에는 전환이 없다고 판단한 후 현재 프레임(c)을 이전 프레임(p)으로 셋팅하고 위의 과정을 다시 수행하면서 검출을 계속한다. 비디오의 모든 프레임들에 걸쳐서 상기 과정을 진행하면 샷 경계 검출과정이 끝나게 된다.

경계 검출을 위해 연속된 모든 프레임의 변화를 측정하는 것은 상당한 계산 시간이 요구된다. 또한 비디오는 시간적 중복성이 크므로 시간적 해상도를 줄여서 모든 프레임이 아닌 매 15프레임마다의 비교를 통해서 경계 검출을 실시한다.

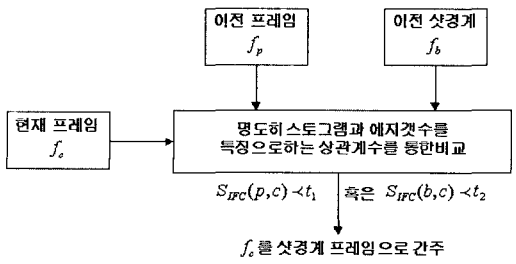


그림 3 샷 경계검출 다이어 그램

3. 특징치 추출

비디오에서 특정 감성을 가진 장면을 검색하기 위해서는 장면의 내용을 효과적으로 표현하는 특징 값을 추출하고 이 특징 값을 이용하여 감성기반 검색을 실현할 수 있어야 한다. 이를 위해 앞에서 설명한 방법을 통해 장면 단위로 샷 경계들을 검출하고 키 프레임을 추출한 후에 키 프레임들의 내용을 표시하는 특징값을 추출하여 장면을 표현하는 특징으로 삼는다. 본 연구에서는 색상 정보, 밝기 정보, 에지 정보 등의 시각적 내용(visual content)을 추출하여 이용하며, 비디오는 특성상 정지영상과는 달리 시간 정보도 포함되기 때문에 장면내의 샷의 지속시간, 전체 샷 변화 중 점진적 변화의 비중등의 시각적 리듬(visual rhythm)을 특징으로 이용한다.

평균 색상 히스토그램(Average Color Histogram)

색상은 비디오의 내용을 표현하는 가장 기본적인 특징이며 전통적으로 내용기반 영상/비디오 검색 등의 여러 연구에서 효과적으로 사용되어 왔다. 이 특징은 특정 감성을 나타내는데도 효과적으로 사용될 수 있다. 예를 들어 따뜻한 영상의 경우는 주로 붉은(red) 성분의 요소가 많고 찬 영상의 경우에는 청색(blue, 혹은 cyan) 성분의 요소가 많다. 비디오의 경우에도 일반적으로 "action" 장면에서는 적색(red), 보라색(purple) 등의 색상이, "quietness" 장면에서는 청색(blue), 녹색(green), 백색(white) 등의 색상이 포함되는 경우가 많다[20]. 따라서 본 연구에서는 샷 내의 키 프레임에서 추출한 RGB 통합 히스토그램(joint histogram) 값을 해당 샷의 전체 프레임 개수로 나타낸 지속시간으로 곱하고, 이와 같은 방법을 장면내의 모든 샷에 대하여 동일하게 적용하고 합산한 후 장면내의 전체 프레임 개수로 나누어 평균 색상 히스토그램을 추출한다.

$$AvgH_{RGB}[i] = \frac{\sum_k H_{RGB}[i] \times ShotLength[k]}{N_T}, \quad i = 0,1,2,...26 \tag{1}$$

여기서, $H_{RGB}[i]$ 는 27차원의 통합 RGB 히스토그램(R, G, B채널을 각각 3개씩 균일하게 양자화하여 통합한 3 ×3 ×3 히스토그램) 중의 i번째 빈 값을, $ShotLength[k]$는 프레임 개수로 표시된 k번째 샷의 길이, N_T는 장면내의 전체 프레임의 개수를 나타낸다.

예를 들어, 장면이 2개의 샷으로 이루어지고 ($k=0,1$), 첫 번째 샷 ($k=0$)은 90개의 프레임으로 ($ShotLength[0]=90$), 두 번째 샷 ($k=1$)은 60개의 프레임으로 ($ShotLength[0]=90$) 구성되어 있고 (즉, 총 150(=90+

60) 개의 프레임으로 *(NT=150)* 구성된 장면), 첫 번째 샷의 키 프레임에서 *i=0*번째 색상의 통합 히스토그램 값이 20이고 *($H_{RGB}[0]=20$)*, 두 번째 샷의 키 프레임에서 *i=0*번째 색상의 통합 히스토그램 값이 30 이라고 *($H_{RGB}[0]=30$)* 가정하면, *i=0* 번째의 평균색상 히스토그램은 *$AvgH_{RGB}[0]=(20\times90+30\times60)/150=24$*가 된다.

평균 밝기(Average Brightness)

밝은 영상은 가벼운 느낌, 행복한 느낌 등을 표현하고 반대로 어두운 영상은 딱딱하고 무거우며 침울한 느낌을 표현한다. 비디오에서도 "quietness" 장면은 주로 밝은 톤의 정보가 많이 포함되는 경우가 많다. 평균 색상 히스토그램과 유사하게 본 연구에서 장면내의 평균 밝기 값을 특징으로 포함시킨다.

$$AvgBright=\frac{\sum_k LocalAvgBrigt[k]\times ShotLength[k]}{N_T} \quad (2)$$

여기서, *$LocalAvgBrigt[k]$* 는 k번째 샷의 키 프레임의 평균 밝기 값을, *$ShotLength[k]$* 는 프레임 개수로 표시된 k번째 샷의 길이, N_T 는 장면내의 전체 프레임의 개수를 나타낸다.

평균 에지 히스토그램(Average Edge Histogram)

우울한 영상은 영상전체가 뭉개진(Blurring) 것같이 주요한 에지의 개수가 적고 유쾌한 분위기의 영상은 상대적으로 반대의 경우가 많다[17]. 본 연구에서는 캐니 에지 검출 연산자를 적용하여 주요한 에지를 구하고 방향별로 에지의 개수를 히스토그램화 하여 특징 값으로 사용한다[7].

$$AvgH_{EDGE}[i]=\frac{\sum_k H_{EDGE}[i]\times ShotLength[k]}{N_T},\quad i=0,1,2,...,71 \quad (3)$$

여기서, *$H_{EDGE}[i]$* 는 72차원의 방향별 에지Histogram중의 i번째 빈 값을, *$ShotLength[k]$* 는 프레임 개수로 표시된 k번째 샷의 길이, N_T 는 장면내의 전체 프레임의 개수를 나타낸다.

평균 샷 시간(Average Shot Duration)

"action", "excitement" 등의 장면은 일반적으로 빠르게 샷이 변화하며 "quietness", "relaxation", "happiness" 등의 장면은 샷의 길이가 길고 샷 내의 변화도 심하지 않다. 본 연구에서는 장면내의 샷의 개수를 구한 후 샷의 지속시간을 평균하여 평균 샷 시간을 감성을 표현하는 특징 값으로 사용한다.

$$AvgShotTime=\frac{\sum_k ShotDuration[k]}{N_S} \quad (4)$$

여기서, *$ShotDuration[k]$* 은 초(second)로 표시된 k번째 샷의 지속시간, Ns 는 장면내의 전체 샷의 개수를 나타낸다.

점진적 샷 변화율(Gradual Change Rate)

비디오에서 점진적 변화에 의한 샷의 경계는 특정한 감성을 불러 일으키는 경우가 있다. "quietness" 장면에서는 디졸브(dissolve)와 같은 점진적 샷의 변화 등을 포함하는 경우가 많다. 따라서 본 연구에서는 장면내의 샷 변화 중에 점진적인 변화가 차지하는 정도를 특징 값으로 이용한다.

$$GradRate=\frac{N_G}{N_S} \quad (5)$$

여기서, N_G 은 점진적 변화에 의한 샷의 개수, N_S 는 장면내의 전체 샷의 개수를 나타낸다.

4. 감성기반 비디오 검색

4.1 대화형 유전자 알고리즘

유전자 알고리즘(Genetic Algorithm, GA)은 자연의 진화이론을 바탕으로 주어진 환경에 잘 적응하는 개체(chromosome)을 선택하고 교차(crossover)하며 때에 따라서는 돌연변이(mutation)도 하여 다음 세대에 우수한 유전 형질을 전달(reproduction)하는 방법으로서 기계학습이나 패턴분류, 최적화문제 등에 효과적으로 사용되고 있다[24]. 본 연구에서는 장면에서 얻을 수 있는 평균 색상 히스토그램(27차원), 평균밝기(1차원), 평균 에지 히스토그램(72차원), 평균 샷 시간(1차원), 점진적 샷 변화율(1차원)의 총 102차원의 특징벡터(그림 4)를 유전자 알고리즘의 개체로 표현하고 사용자의 감성에 적합하도록 진화시킴으로써 감성기반 검색을 달성한다.

대화형 유전자 알고리즘(Interactive Genetic Algorithm)은 목적함수가 명시적으로 정의되지 않을 경우 적합도 함수로 사람의 판단(human evaluation)을 채택하는 기술이다. 이러한 특성은 사람의 주관적 평가에 기반한 시스템을 개발하는데 적합하다. 예를 들면 디자인 영역이나 작곡과 같이 인간의 기호를 반영하며 최적화를 요구하는 시스템의 개발 시 개인의 판단 이외에는 이 시스템의 성능을 평가할 측정법이 존재하지 않는다. 이러한 경우 대화형 유전자 알고리즘은 사람에게 심리적인 공간(psychological space)상에서의 잠재적인 목표(potential target)와 실제 시스템에서 생성한 결과 사이의 차이를 평가하게 하면서 유전자 알고리즘으로 파라미터 공간(parameter space)을 탐색한다. 즉, 대화형 유전자 알고리즘은 사람과 유전자 알고리즘의 상호협동을 통하여 두 공간상의 대응관계에 기반한 최적화된 시스템의 개발을 가능하게 하는 기술이다. 이러한 특성으로 인해 그래픽 아트, 산업 디자인, 작곡과 같은 예술분야,

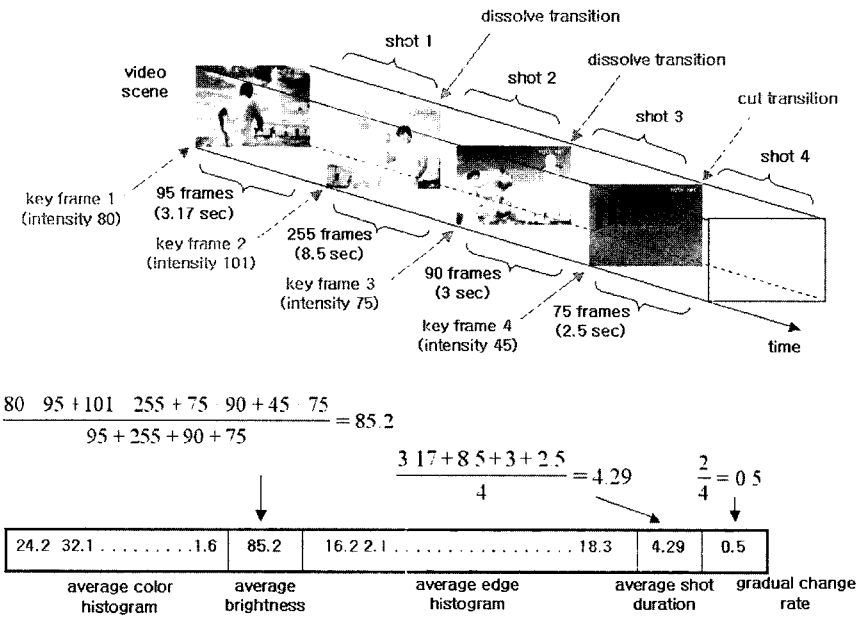


그림 4 유전자 개체(chromosome) 표현

음성처리, 가상현실, 정보검색, 그리고 교육, 게임 등 다양한 분야에서 폭넓게 적용되고 있다[17,18,25-28]. 대화형 유전자 알고리즘을 이용한 광범위한 참고문헌과 적용 분야는 Tagaki의 논문[29]에서 찾아 볼 수 있다. 본 연구에서 제안하고자 하는 장면 단위의 감성기반 비디오 검색의 경우도 특정 감성을 내포하고 있는 장면에 대한 판단을 인간이 할 수 밖에 없으므로 대화형 유전자 알고리즘을 이용하여 검색을 실시 하고자 한다.

4.2 검색 방법

장면단위 검색을 위해서는 여러 장면으로 이루어진 비디오를 장면단위로 분할해서 데이타베이스를 구성하거나, 특정한 장면만을 담은 짧은 비디오(예를 들면 CF 같은 상업광고)를 가지고 데이타베이스를 구성해야 한다. 본 연구에서는 실험의 편의를 위해 후자의 경우를 사용했다. 이 들 비디오를 사용하여 감성에 기반한 검색은 아래와 같은 절차에 의해 수행된다.

Step 1. 특징을 추출하여 유전자 개체로 표현하고 데이타베이스에 색인한다.

데이타베이스에서 한 개의 비디오를 선택하여 샷 경계검출 모듈에 입력하면 모듈에 의해 급진적 경계(cut)와 점진적 경계(dissolve등)에 해당하는 샷을 검출한다. 이 때 검출된 각 샷들의 시작 프레임을 해당 샷의 키 프레임으로 선택했다. 다음에는 각 키 프레임들을 특징추출 모듈에 입력하여 평균 색상 히스토그램, 평균밝기, 평균 에지 히스토그램, 평균 샷 시간, 점진적 샷 변화율의 특징들을 추출하고 데이타 베이스와 검색테이블에 입력한다. 따라서 한 개의 비디오에서 특징을 나타내는 한 개의 유전자 개체가 생성된다(그림 4참조). 이와 같은 과정을 데이타베이스에 저장된 모든 비디오에 대해 실시하여 비디오 개수와 동일한 유전자개체를 생성하는 특징추출과정을 마친다(그림 1의 상부 네모의 *indexing* 부분).

Step 2. 초기 비디오 집단을 제시하고 검색하고자 하는 감성비디오를 선택한다.

시스템은 임의의 비디오 15개를 제시한 후 사용자로부터 검색하고자 하는 감성을 표현하는 비디오가 있는지를 선택하도록 한다. 비디오를 제시하는 방법은 각 비디오의 첫 번째 프레임을 보여주고, 각 비디오 밑에 있는 재생 버튼을 눌러 해당 비디오를 보면서 사용자가 원하는 감성을 반영하고 있는지를 조사하여 반영하고 있으면 선택하고 그렇지 않으면 선택하지 않는다(그림 5의 시스템 GUI를 참고). 때로는 사용자에 따라 비디오를 재생하면서 보는 것이 시간상 부담스러울 수도 있기 때문에 각 비디오의 키 프레임만을 보고도 선택할 수 있도록 해당 비디오 밑에 키 프레임을 보여주는 기능도 삽입했다(그림 6).

Step 3. 선택된 비디오집단에 교차 유전연산을 적용한다.

선택된 각각의 비디오에 해당하는 염색체들을 데이타베이스에서 추출하여, 이들 간에 교차 유전 연산자를 적용하여 새로운 염색체의 집단을 생성한다. 교차 연산을 적용하는 방법은 먼저 선택된 비디오들에서 임의로 쌍

들을 선택하고, 각 쌍들에 대해서 연산을 적용하기 위한 위치를 임의로 선택한다. 본 연구에서는 4개의 교차점 중 1개의 위치만을 선택하여 연산을 수행한다(그림 7).

Step 4. 유전연산후의 목표 비디오 개체와 데이타베이스간의 유사도를 계산한다.

연산을 수행한 후에는 해당 감성을 적절히 표현하는 15개의 새로운 염색체가 형성되며 이 15개의 새로운 목표 비디오 염색체와 가장 유사한 비디오 15개를 유사도 검색 함수 S(Q,D)에 의해 데이타베이스에서 찾아 원하는 감성을 표현하는 다음세대의 집단으로 제시한다. 유사도는 염색체 정보를 기반으로 특징들간의 유클리드 거리를 구하고 최종적으로 모든 거리를 통합하여 계산한다.

$$S(Q,D) = w_{RGB} \times S_{RGB}(Q,D) + w_{BRIGHT} \times S_{BRIGHT}(Q,D) + w_{EDGE} \times S_{EDGE}(Q,D) + w_{SHOTTIME} \times S_{SHOTTIME}(Q,D) + w_{GRADRATE} \times S_{GRADRATE}(Q,D)$$

$$S_{RGB}(Q,D) = \sum_{i=0}^{31} \left(AvgH^{Q}_{RGB}[i] - AvgH^{D}_{RGB}[i]\right)^{1/2}$$

$$S_{BRIGHT}(Q,D) = \left(AvgBright_{Q} - AvgBright_{D}\right)$$

$$S_{EDGE}(Q,D) = \sum_{i=0}^{71} \left(AvgH^{Q}_{EDGE}[i] - AvgH^{D}_{EDGE}[i]\right)^{1/2}$$

$$S_{SHOTTIME}(Q,D) = \left(AvgShotTime_{Q} - AvgShotTime_{D}\right)^{1/2}$$

$$S_{GRADRATE}(Q,D) = \left(GradRate_{Q} - \text{Gradrate}_{D}\right)^{1/2} \quad (6)$$

여기서, $S_{RGB}(Q,D)$는 질의 비디오와 데이타베이스 비디오간의 평균 색상 히스토그램을 이용한 유사도 값을, $S_{BRIGHT}(Q,D)$는 평균 밝기를 이용한 유사도 값을, $S_{EDGE}(Q,D)$는 평균 에지 히스토그램을 이용한 유사도 값을, $S_{SHOTTIME}(Q,D)$는 평균 샷 시간을 이용한 유사도 값을, $S_{GRADRATE}(Q,D)$는 점진적 샷 변화율을 이용한 유사도 값을 나타낸다. ω_{RGB}, ω_{BRIGHT}, ω_{EDGE}, $\omega_{SHOTTIME}$, $\omega_{GRADRATE}$는 $\omega_{RGB} + \omega_{BRIGHT} + \omega_{EDGE} + \omega_{SHOTTIME} + \omega_{GRADRATE} = 1$을 만족하는 각각의 유사도에 대한 가중치를 나타낸다.

$S(Q,D)$는 값이 작을수록 유사도가 높은 것을 의미한다. 실험에서는 5개 각각의 유사도를 겹합하기 위해서 [0,1]범위로 정규화를 하였으며, 가중치 값은 $\omega_{RGB} = 0.2$, $\omega_{BRIGHT} = 0.2$, $\omega_{EDGE} = 0.1$, $\omega_{SHOTTIME} = 0.25$, $\omega_{GRADRATE} = 0.25$를 디폴트로 하였다. 이 초기 가중치는 여러 번의 실험을 거쳐 경험에 의해 설정된 것이며, 또한 시각적 내용($\omega_{RGB} = 0.2$, $\omega_{BRIGHT} = 0.2$, $\omega_{EDGE} = 0.1$)에 0.5, 시각적 리듬($\omega_{SHOTTIME} = 0.25$, $\omega_{GRADRATE} = 0.25$)에 0.5로 각각 균등하게 설정되었다.

Step 5. 종료조건을 만족할 때까지 Step2 - Step4를 반복한다.

Step2에서 Step4까지를 종료조건이 만족할 때까지 여러 번 반복함으로써 이 후 세대에서 원하는 감성의 비디오를 검색하게 된다. 종료조건은 현재 단계에서 검색결과가 만족스럽거나 더 많은 반복에도 결과가 변화가 없을 때, 또는 사용자의 피곤함 때문에 더 이상 수행이 어려울 때이다.

5. 실험

5.1 시스템 환경

이 시스템은 펜티엄 PC에서 Visual C++를 이용해 구현하였으며, 300개의 광고 동영상 (총 2.5 Giga Byte)을 데이타베이스로 사용하였다. 광고 동영상을 사용한 이유는 대부분의 동영상이 1분 이내의 짧은 시간동안 동일한 한가지 주제로 이루어 지는 경우가 많으므로 본 연구에서 제안한 장면 단위 감성분석에 적합하기 때문이다. 교차율은 0.54를 적용하였다. 즉, 15개의 초기 비디오 집단에서 약 8개의 비디오가 유전자 교차를 위해 선택된다.

그림 5의 GUI에서 최초 15개의 초기 비디오 집단은 데이타베이스 내에서 임의로 선택된 동영상이다. 이 중에서 사용자가 각 동영상을 재생시키거나 키 프레임만을 보면서 자신이 찾기 원하는 감성을 가진 분위기의 동영상을 선택하면 선택된 비디오를 기반으로 8개의 교차연산을 위한 비디오가 선택되고 1점 교차연산에 의해 8개의 새로운 염색체를 포함하여 총 15개의 염색체가 생성된다. 다음에는 각각의 염색체와 가장 유사한 비디오를 데이타베이스에서 유사도함수에 의해 계산하여 다음 세대의 비디오를 보여준다. 이 과정을 사용자가 원하는 비디오를 찾을 때 까지 반복한다. 또한 검색과정 중에 필요에 따라 사용자가 5개의 유사도 가중치를 동적으로 변환시킬 수 있도록 하였다.

5.2 실험 및 분석

실험에 사용된 감성은 "action", "excitement", "suspense", "quietness", "relaxation", "happiness" 의 6가지 이다. 그러나 본 알고리즘은 사용된 6개의 감정에 초점을 맞추어서 제안된 것이 아니며 얼마든지 다른 감성에도 적용할 수 있다. 여기서 사용된 감성은 [20]에서 상업용 광고 비디오의 감성을 표현하는데도 사용되었다. [20]에 따르면 "action" 장면은 붉은 계통의 색상을 띄면서 움직임 정보가 많고 짧은 시간동안 여러 개의 컷들로 연결된 경우가 많고, "excitement"도 짧은 시간 동안 여러 개의 컷들로 연결된 경우가 많으며, "suspense" 장면은 "action"의 특징과 더불어 컷들로 연결된 여러 개의 길고 짧은 시퀀스로 이루어진 경우가 많다. "quietness"는 청색, 녹색, 흰색 등의 색상을 가지면서 긴 시간동안의 점진적 경계인 디졸브(dissolve) 등으로 이루어진 경우가 많고, "relaxation"은 보통 움직임 정보가 없으며, "happiness"는 "quietness" 장면에 약간

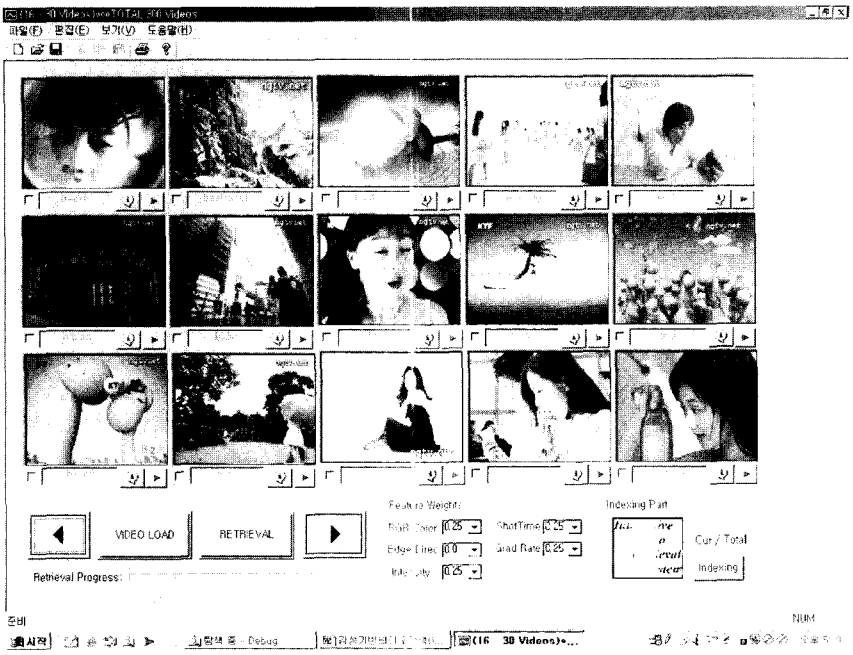
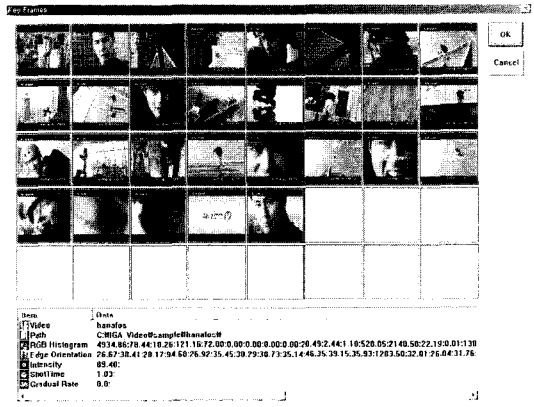


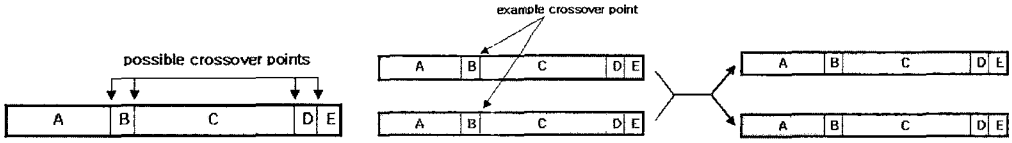
그림 5 감성기반 비디오 검색 시스템의 GUI (Graphic User Interface)

(a) 비디오를 재생해서 보는 방법(그림 5의 GUI에서 영상 밑에 있는 ▶을 누를 경우에 실행됨)



(b) 키 프레임만 보는 방법(그림 5의 GUI에서 영상 밑에 있는 ⓘ를 누를 경우에 실행됨)

그림 6 비디오의 감성을 판단하는 인터페이스



(a) 가능한 교차점들 (b) 1점 교차의 예

그림 7 교차 유전자 연산

의 움직임 정보 등이 포함되는 경우가 많다. 이상과 같은 특징은 본 논문에서 제안된 5가지 특징을 연결한 염색체로 나타낼 수 있다. 10명의 피험자들에 의해 자의적 판단으로 300개의 비디오를 분류한 결과 "action"은 39

개, "excitement"는 56개, "suspense"는 22개, "quietness"는 72개, "relaxation"은 71개, "happiness"는 82개 였다. 그림 8은 위의 6가지 감성을 가진 비디오의 키 프레임을 나열한 것이다.

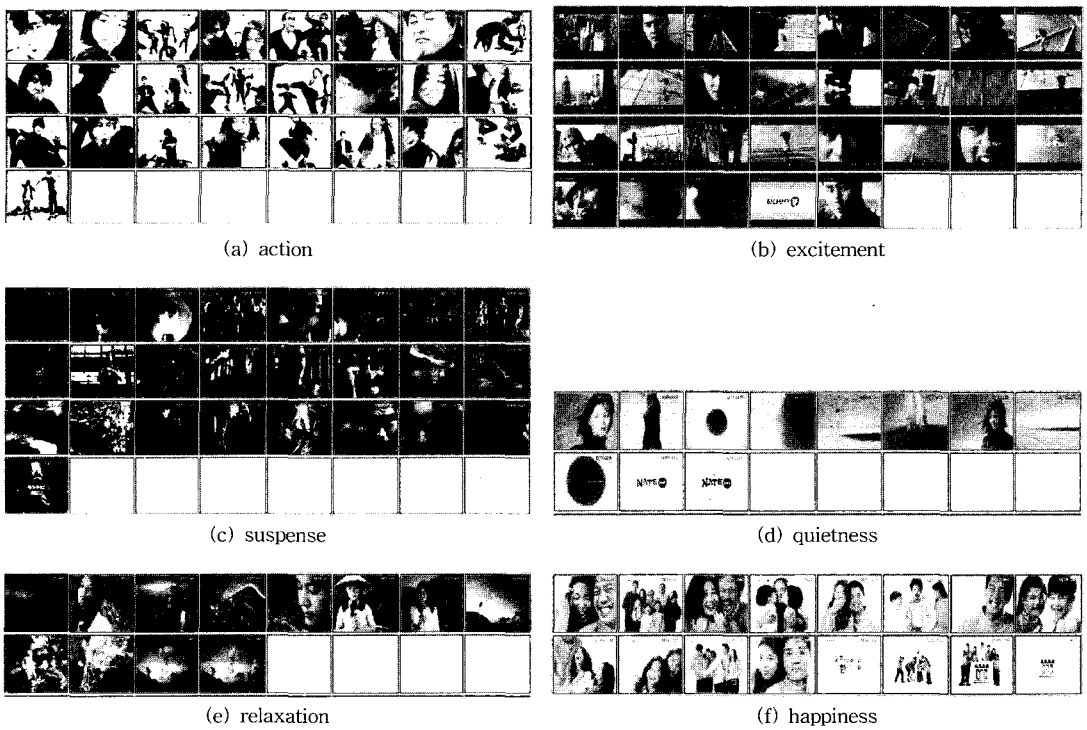
알고리즘의 성능 평가를 위해 2가지 실험을 하였다. 먼저 염색체로 표현된 평균 색상 히스토그램, 평균 밝기, 평균 에지 히스토그램, 평균 샷 시간, 점진적 샷 변화율의 5가지 특징이 대화형 유전자 알고리즘의 유전자 표현으로 유용한지를 알아보기 위한 실험을 하였다. 둘째로, 알고리즘의 성능 평가를 위해 최대 10세대까지 검색을 실시한 후 사용자 입장에서 검색에 대한 만족도(시스템의 효과성)를 알아 보았다.

5.2.1 특징 벡터의 유용성

300개의 데이타 베이스에서 피험자들에 의해 6가지의 감성으로 분류되어 군집화 된 비디오에서, 군집내의 비디오간의 유사도(intra-class similarities)와 군집간 비디오의 유사도(inter-class similarities)를 유사도 계산식 (6)에 의해 구했다. 실험의 편의를 위해 각 감성에 해당하는 비디오 중에서 30%의 비디오만을 임의적으로 선택하여 실험하였다("action 12개, excitement 17개, "suspense" 7개, "quietness" 22개, "relaxation" 21개, "happiness" 25개).

결과적으로 이 값은 같은 부류의 비디오가 그렇지 않은 비디오에 비해 특징 공간상에서 거리가 가까운지를 판단하는 근거가 된다. 각 감성에 대한 군집내의 유사도와 군집간 유사도를 그림 9에 박스 그림(box plot)으로 나타냈다. 여기에서 박스의 크기는 유사도 값의 50%의 데이타를 나타낸다. 위쪽 수평선은 유사도 값의 75% 퀀타일(quantile)을 나타내고 하위 수평선은 유사도 값의 25% 퀀타일을 나타낸다. 박스 내의 수평선은 중간값(median)을 나타낸다. 박스 외곽의 수평선은 각각 최대 유사도와 최소 유사도를 나타낸다. 표시된 수치는 값이 작을수록 유사도가 높은 것이고, 유사도가 높다는 것은 두 비디오가 특징 공간상에서 가까운 위치에 있다는 것을 의미한다. 예를 들어 그림 9(a) "action"내의 12개 각각의 비디오를 질의로 하여 유사도를 검색했을 때 "action"비디오들 간의 유사도 값들(군집내 유사도 값들)을 나타내는 "action-action"이 12개의 "action" 비디오의 질의에 대한 "quietness"내의 비디오들의 유사도 값들(군집간 유사도 값들)을 나타내는 4번째 박스의 "action-quietness"보다 전체적으로 낮은 값을 가지므로 유사도가 높은 것을 의미한다.

그림 9(a)에서 "action-action"이 "action-excitement", "action-quietness", "action-relaxation", "action-happi-



(a) action (b) excitement

(c) suspense (d) quietness

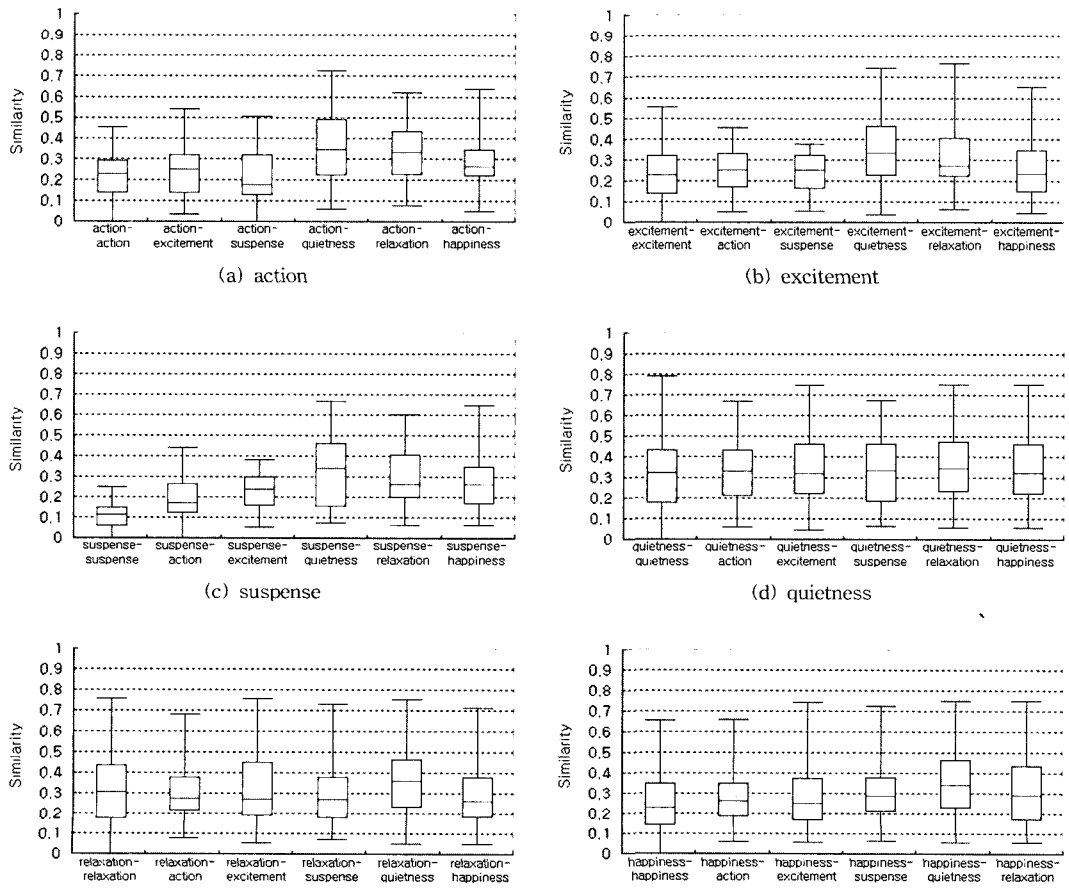
(e) relaxation (f) happiness

그림 8 6가지 감성을 가진 비디오의 키 프레임 예

ness"보다 박스와 중간값이 아래에 위치하고 최대, 최소값도 아래에 위치하므로 보다 높은 유사도를 가지고 있음을 알 수 있다. 그러나 3번째 "action-suspense"의 경우에는 최대값은 낮지만 중간값이 높고, 또한 25% 퀀타일은 높고 75% 퀀타일은 낮아서 어느 것이 높은 유사도를 가지고 있는지 판단하기가 어렵다. 그림 9(b)에서도 "excitement-excitement"는 "excitement-quietness", "excitement-relaxation", "excitement-happiness"보다 높은 유사도를 가지고 있음을 알 수 있고, "excitement-action", "excitement-suspense"등과는 뚜렷한 차이점을 알 수 없다. 그림 9(c)에서는 군집내 유사도 "suspense-suspense"가 군집간 유사도에 비해 가장 높은 유사도를 가지고 있음을 알 수 있다. 그림 9(d)의 "quietness"와 그림 9(e)의 "relaxation"에서는 군집내의 유사도와 군집간의 유사도의 차이점을 발견하기 어렵다. 그림 9(e) "happiness"는 군집내 유사도가 군집간 유사도보다 높음을 알 수 있다. 따라서 전체적으로 판단이 어려운 "quietness"와 "relaxation"을 제외한 감성 대부분이 같은 군집에 속하는 비디오가 다른 군집에 속하는 비디오보다 전반적으로 유사도가 높음을 알 수 있다. 따라서 본 논문에서 사용한 특징들이 6가지 감성을 어느 정도 적절히 표현한다고 말할 수 있다.

5.2.2 시스템의 효과성

알고리즘의 성능 평가를 위해 사용자 입장에서 검색에 대한 만족도(시스템의 효과성)를 평가하였다. 이를 위해 최대 10세대까지의 검색된 비디오에 대한 만족도를 알아 보았다. 실험은 10명의 대학원생에게 시스템의 전반적인 사용방법과 각 특징들의 의미를 설명한 후 수행하도록 하였다. 그리고 객관적인 결과를 얻기 위해 각 실험자 마다 동일한 주제에 대해 이전과 다른 초기 15개의 비디오 집단을 제시한 후 실험을 하도록 요구하였고 이렇게 해서 얻어진 만족도의 평균을 각 감성에 대한 검색의 만족도로 사용하였다.



(a) action (b) excitement

(c) suspense (d) quietness

(e) relaxation (f) happiness

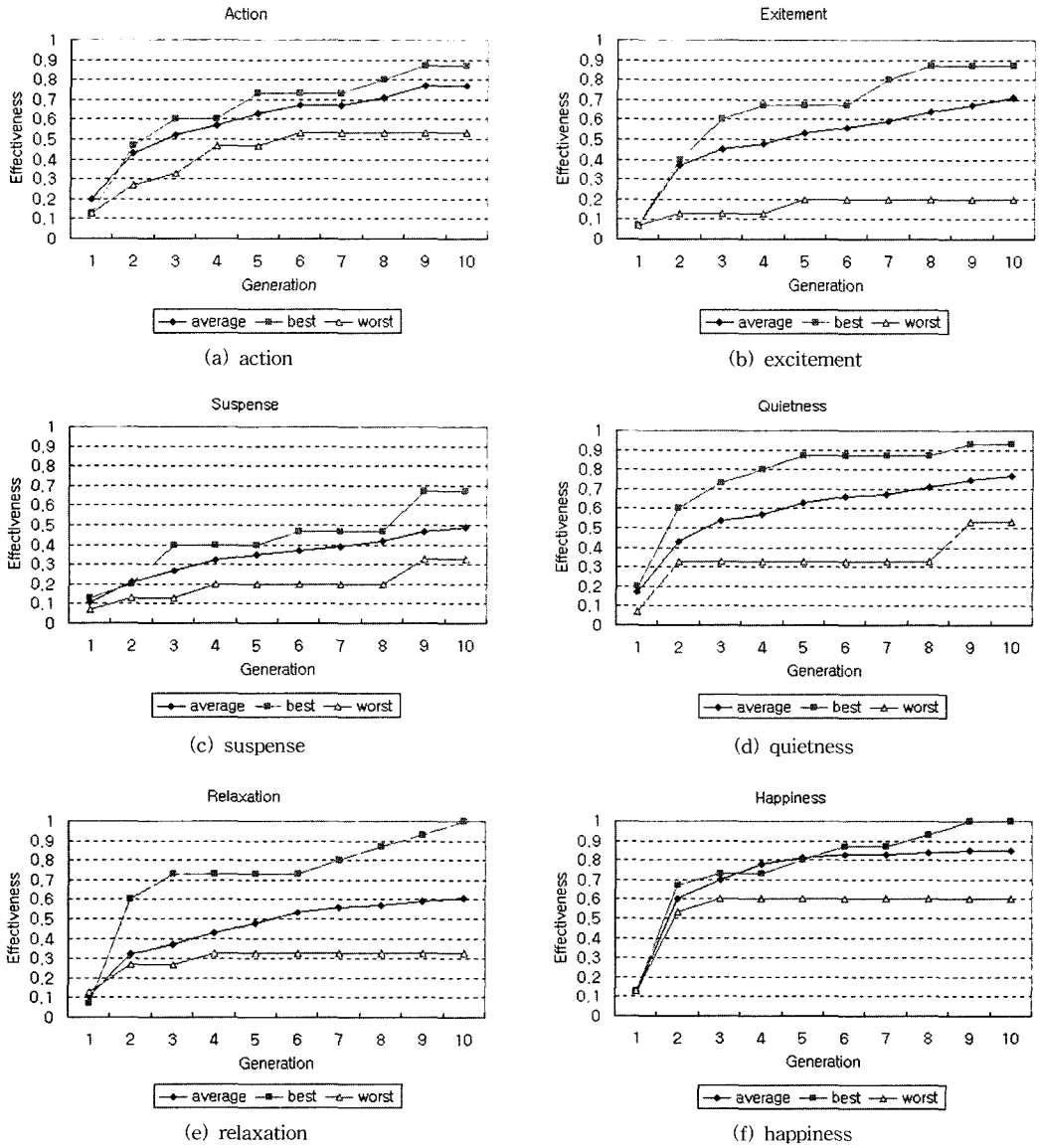
그림 9 감성에 대한 군집내 유사도 값(맨 좌측 박스)과 군집간 유사도 값

$$Effectiveness = \frac{N_{CORRECT}}{N_{TOTAL}} \quad (7)$$

여기서 N_{TOTAL}은 비디오 집단 수 15이고 $N_{CORRECT}$는 질의 감성을 만족하는 비디오 개수의 평균을 나타낸다.

그림 10은 10세대까지의 6가지 감성에 대한 효과성의 추이를 3가지의 경우로 나누어 가장 좋은 결과를 보일 때, 가장 나쁜 결과를 보일 때, 그리고 모든 결과를 평균한 것을 나타낸다. 모든 감성에 걸쳐서 세대가 증가할수록 만족도가 증가함을 볼 수 있으며 10세대 후에는 평균적으로 0.7정도의 만족도를 보여준다(즉, 10세대 후에는 총 15개의 비디오 중에 10개 이상의 만족된 비디오를 얻을 수 있다). 전체적으로 "suspense"와 "relaxation"을 제외하고는 모두 0.7이상의 높은 만족도를 나타냈다. "happiness"의 경우는 가장 높은 0.85의 만족도를 보였으며, "suspense"의 경우는 가장 낮은 0.49의 만족도를 보였다(표 2).

"action"의 경우 10세대 후에는 최대 0.87의 만족도와, 최소 0.53의 만족도, 평균 0.77의 만족도를 얻었다. 실험을 통해서 짧은 시간동안 여러 개의 새로운 컷으로



(a) action (b) excitement

(c) suspense (d) quietness

(e) relaxation (f) happiness

그림 10 각 감성에 대한 사용자의 만족도(시스템의 효과성)

이루어진 샷이 많았기 때문에 시간적인 정보가 중요함을 느낄 수 있었다.

"excitement"의 경우 10세대 후에는 최대 0.87의 만족도와, 최소 0.2의 만족도, 평균 0.71의 만족도를 얻었다. 0.2의 저조한 만족도는 초기 비디오 집단에서 "excitement"관련 비디오가 일반 적으로"excitement" 감성에서 보여지는 것과 달리 평균 샷 시간이 길고 디졸브로 표현된 점진적 장면전환이 포함되었기 때문이다.

"suspense"의 경우 10세대 후에는 최대 0.67의 만족도와, 최소 0.33의 만족도, 평균 0.49의 만족도를 얻었다. 이 결과치는 실험에 사용된 6가지 감성에 대해서 가장 낮은 만족도이다. 염색체로 사용된 5가지의 특징이 해당 감성을 충분히 표현하지 못한 것으로 생각된다.

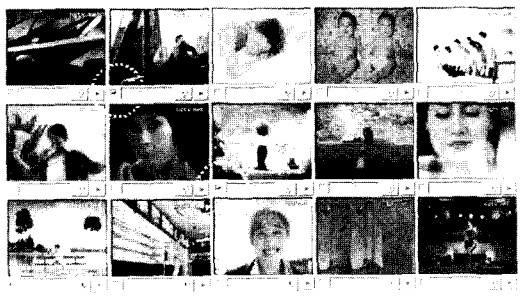
"quietness"의 경우 10세대 후에는 최대 0.93의 만족도와, 최소 0.53의 만족도, 평균 0.77의 만족도를 얻었다. 실험을 통해서 긴 시간동안 여러 개의 새로운 컷이나 디졸브로 이루어진 샷이 많았기 때문에 시간적인 정보가 중요함을 느낄 수 있었다.

"relaxation"의 경우 10세대 후에는 최대 1.0의 만족도와, 최소 0.33의 만족도, 평균 0.61의 만족도를 얻었다. 실험을 통해서 긴 시간동안 여러 개의 새로운 컷, 혹은 디졸브로 이루어진 샷이 많았기 때문에 시간적인 정보가 중요함을 느낄 수 있었다.

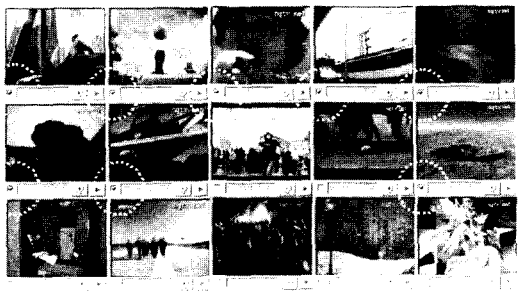
"happiness"의 경우 10세대 후에는 최대 1.0의 만족도와, 최소 0.6의 만족도, 평균 0.85의 만족도를 얻었다. 이 결과치는 실험에 사용된 6가지 감성에 대해서 가장 높은 만족도이다. 300개의 비디오 중에 82개의 비디오가 "happiness"와 관련이 있었으며 대부분이 흰색, 녹색, 청색 등으로 이루어 지면서 긴 평균 샷 시간을 가지고 있었다. "happiness"의 경우에는 동적 가중치 변화를 적게 하였다.

그림 10을 보면 전체적으로 증가추세에 있으므로 10세대 이후에 보다 좋은 결과를 얻을 수 있는 가능성이 보인다. 그러나 이는 시간적인 제약과 사용자의 피로를 가중 시킬 수 있다는 단점이 있다. 따라서 이러한 단점에도 불구하고 보다 높은 만족도를 얻기 위해서는 여러 세대에 걸쳐서 계속적인 검색을 시도해 볼 수도 있을 것이다.

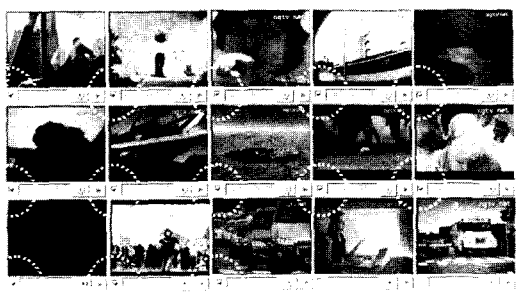
그림 11은 "action"에 대해 세대가 진행됨에 따라 어떤 비디오들이 검색되는지를 보여준다. 그림 11(a)는 최초에(1세대) 임의적으로 선택된 15개의 비디오 집단인데, 2개의 적합한(사용자가 체크한 dblee와 doublichng를 참조) 비디오가 선택되어 있음을 알 수 있다. 그림 11(b)와 11(c)는 대화형 유전자 알고리즘에 의해 사용자 피드백을 실시한 후에 각각 5세대와 10세대 후의 검색된 비디오 집단을 보여준다. 그림에서 보듯이 5세대 후



(a) 1세대의 비디오(2개가 적합한 것으로 체크됨)



(b) 5세대 후의 비디오(8개가 적합)



(c) 10세대 후의 비디오(13개가 적합)

그림 11 "action"에 대해 세대가 진행됨에 따라 검색된 결과 비디오

에는 8개의 적합한 비디오를 얻을 수 있었고 10세대 후에는 13개의 적합한 비디오를 얻을 수 있었다.

또한 사용자에 의한 적합도 선택이 아닌 임의의 선택(random selection)의 경우에 만족도의 추이가 어떻게 되는지를 알아보았다(그림 12참조). 임의 선택은 *random*함수를 사용하여 임의의 비디오를 임의의 개수만큼 선택하고, 또 선택된 비디오들간에 교차연산의 쌍을 임의로 선택하였다. 짐작한데로 그림 10과 비교하여 모든 감성에 대해서 낮은 만족도를 보였으며 세대가 진행함에 따라 만족도의 증가추세도 보이지 않았다.

본 연구에서 사용된 상업용 광고 동영상은 짧은 시간 동안에 많은 내용을 소비자에게 전달해야 하기 때문에

표 2 10세대후의 검색결과 비디오에 대한 평균적인 만족도

	action	excitement	suspense	quietness	relaxation	happiness
효과성	0.77	0.71	0.49	0.77	0.61	0.85
평균	0.7					

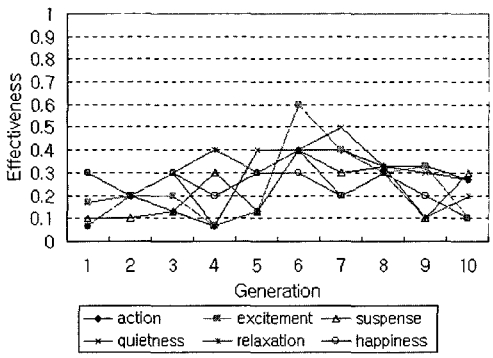


그림 12 임의선택(random selection)을 통한 각 감성에 대한 사용자의 만족도

내용의 변화가 매우 급격하다. 그러나 뉴스, 드라마, 영화 등의 일반 비디오의 경우는 상대적으로 변화가 급격하지 않다. 따라서 본 연구에서 제안한 방법을 일반적인 비디오의 장면에 적용할 경우는(예를 들어 뉴스의 앵커 장면, 드라마의 대화장면 등) 보다 높은 검색 만족도를 얻을 수 있을 것으로 생각된다. 또한 주목하여야 할 점은 본 연구에서 비디오의 적합성을 적합/적합하지 않음의 2가지 경우만을 체크박스를 통해 나타냈지만 [17]에서는 감성기반 영상검색을 위해 슬라이더바를 이용하여 감성의 정도를 0-7로 분할하여 적용하였다. 이와 같이 할 경우 보다 적합도가 높은 영상들이 교차후보로 선택될 가능성이 높기 때문에 다음세대에 보다 높은 만족도를 가진 해를 얻을 수 있을 것이다. 그러나 본 연구에서는 이와 같은 방법은 사용자의 피로를 증가시킬 우려가 있기 때문에 사용하지 않았다.

6. 결 론

본 논문에서는 감성에 기반한 비디오 검색방법을 제안하였다. 대화형 유전자 알고리즘을 이용하여 사용자가 막연하게 가지고 있는 감성공간과 물리적 특징공간과의 매핑을 통하여 감성기반 비디오 검색을 실현하였다. 제안된 방법은 사용자가 제시된 비디오들 중 자신이 찾고자 원하는 감성을 내포하는 비디오를 선택하면 교차 유전연산과 유사도 비교에 의해 새로운 후보 비디오가 제시되고, 이렇게 비디오 제시와 선택의 과정이 사용자가 원하는 목표 비디오를 찾을 때까지 계속된다.

300개의 상업용 광고 비디오에 대해 "action", "excitement", "suspense", "quietness", "relaxation", "happiness" 감정을 표현하는 검색을 수행한 결과 10세대 후에 평균 70%의 만족도를 얻을 수 있었고, 이 후 세대에 계속적인 검색을 실시할 경우 보다 높은 만족도를 얻을 수 있는 가능성을 제시하였다.

제안된 방법이 감성기반 비디오 검색에 대한 새로운 방법을 제시하였지만, 몇 가지 해결해야 할 문제점들이 남아 있다. 먼저 비디오로부터 추출된 5가지의 특징인 "평균 색상 히스토그램", "평균 밝기", "평균 에지 히스토그램", 평균 샷 시간", "점진적 샷 변화율"은 비디오 내에서 움직이는 객체 단위의 이동으로 생기는 감성은 제대로 나타낼 수 없다. 따라서 비디오 내의 움직이는 객체를 분리하고 이동벡터나 광류 정보(optical flow) 등을 이용하여 보다 적합한 특징들로 염색체를 표현할 필요가 있다. 또한 대화형 유전자 알고리즘의 가장 큰 문제는 평가자로서의 사람이 느끼는 피곤함이 크기 때문에 많은 세대를 거쳐 진화 시킬 수 없고 작은 집단 내에서 탐색이 이루어져야 한다는 점이다. 이 문제점을 해결하기 위해 비디오를 재생하는 방법과 키 프레임단위로 빠르게 비디오를 파악하는 방법을 제시하였지만, 아무래도 여러 세대에 걸쳐서 검색을 수행하다 보면 사용자의 피곤함 때문에 좋은 검색결과를 얻지 못할 경우도 생길 것이다. 따라서 입력 인터페이스와 디스플레이 인터페이스를 향상시키는 법, 유전자 알고리즘의 수렴을 가속화 시키는 방법과 같은 여러 가지 개선 방안이 향후 연구과제로 남는다. 또한, 알고리즘의 실행시간 측면에서는 질의 감성을 표현하는 15개의 염색체에 대해 가장 유사한 15개의 염색체를 검색해 내는 것이므로 일반적인 비디오 검색 방법보다는 효율성이 떨어지는 문제점이 있다.

참 고 문 헌

[1] M. Flickner, H. Sawhney, W. Niblack, J.Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by Image Content: The QBIC System," *IEEE Computer*, vol. 28, no. 9, pp. 23-31, 1995.

[2] A. Pentland, R.W. Picard, and S. Sclaroff, "Photobook: Content-Based Manipulation of Image Databases," *International Journal of Computer Vision*, vol. 18, no. 3, pp. 233-254, 1996.

[3] J.R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R.C. Jain, and C. Shu, "The Virage Image Search Engine: An Open

Framework for Image Management," In Proc. *SPIE Vol. 2670: Storage and Retrieval for Images and Video Databases IV*, pp. 76-86, 1996.
[4] J.R. Smith and S.-E. Chang, "VisualSEEK: A Fully Automated Content-Based Image Query System," in Proc. *ACM Multimedia*, pp.87-98, 1996.
[5] W.Y. Ma and B.S. Manjunath, "Netra: A Toolbox for Navigating Large Image Databases," *Multimedia Systems*, vol. 7, no. 3, pp. 184-198, 1999.
[6] C. Carson, S. Belongie, H. Greenspan, and J. Malick, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026-1038, 2002.
[7] H.-W. Yoo, D.-S. Jang, S.-H. Jung, J.-H. Park, and K.-S. Song, "Visual Information Retrieval System via Content-Based Approach," *Pattern Recognition*, vol. 35, no. 3, pp. 749-769, 2002.
[8] H.-W. Yoo, S.-H. Jung, D.-S. Jang, and Y.-K. Na, "Extraction of Major Object Features Using VQ Clustering for Content-Based Image Retrieval," *Pattern Recognition*, vol. 35, no. 5, pp. 1115-1126, 2002.
[9] B. T. Truong, C. Dorai, and S. Venkatesh, "New Enhancements to Cut, Fade, and Dissolve Detection Processes in Video Segmentation," in Proc. *ACM Int. Conf. on Multimedia*, pp.219-227, 2000.
[10] U. Gargi, R. Kasturi, and S. H. Strayer, "Performance Characterization of Video-Shot-Change Detection Methods," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 1, pp. 1- 13, 2000.
[11] T. P. Minka and R. W. Picard, "Interactive Learning Using a Society of Models," *Pattern Recognition*, vol. 30, no.3, pp. 565-581, 1997.
[12] A. Vailaya, A. K. Jain, and H.J Zhang, "On Image Classification: City Images vs. Landscapes," *Pattern Recognition*, vol. 31, no. 12, pp. 1921-1936, 1998.
[13] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.J Zhang, "Image Classification for Content-based Indexing," *IEEE Trans. on Image Processing*, vol. 10, no. 1, pp. 117-130, 2001.
[14] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrota, "Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval," *IEEE Trans. on Circuits and Systems Video Technology*, vol. 8, no. 5, pp. 644-655, 1998.
[15] I.J. Cox, M.L. Miller, T.P. Minka, T.V. Papathomas, and P.N. Yianilos, "The Bayesian Image Retrieval System, PicHunter : Theory, Implementation and Psycophysical Experiments," *IEEE Trans. on Image Processing*, vol. 9, no 1, pp. 20-37, 2000.
[16] T. Soen, T. Shimada, and M. Akita, "Objective Evaluation of Color Design," *Color Research and Application*, vol. 12, no. 4, pp.184-194, 1987.
[17] S.-B. Cho, "Towards Creative Evolutionary Systems with Interactive Genetic Algorithm," *Applied Intelligence*, vol. 16, no. 2, pp. 129-138, 2002.
[18] H. Takagi, T. Noda, and S-B. Cho, "Psychological Space to Hold Impression among Media in Common for Media Database Retrieval System," in Proc. *IEEE Int. Conf. on System, Man, and Cybernetics*, pp.263-268, 1999.
[19] J.-S. Um, K.-B. Eum, and J.-W. Lee, "A Study of the Emotional Evaluation Models of Color Patterns Based on the Adaptive Fuzzy System and the Neural Network," *Color Research and Application*, vol. 27, no. 3, pp. 208-216, 2002.
[20] C. Colombo, A. Del Bimbo, and P. Pala, "Semantics in Visual Information Retrieval," *IEEE Multimedia*, vol. 6, no. 3, pp.38-53, 1999.
[21] C. Colombo, A. Del Bimbo, and P. Pala, "Retrieval of Commercials by Semantic Content: The Semiotic Perspective," *Multimedia Tools and Applications*, vol. 13, no. 1, pp. 93-118, 2001.
[22] J. Itten, Art of Color (Kunst der Farbe), Otto Maier Verlag, Ravensburg, Germany, 1961 (in German).
[23] H.-W. Yoo and D-S. Jang, "Automated Video Segmentation Using Computer Vision Technique," *International Journal of Information Technology and Decision Making*, vol. 2, no. 4, 2003 (To appear).
[24] D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989.
[25] J. A. Biles, "GenJam: A Genetic Algorithm for Generating Jazz Solos," in Proc. *Int. Computer Music Conf.*, pp. 131-137, 1994.
[26] C. Caldwell and V. S. Johnston, "Tracking a Criminal Suspect through Face-Space with a Genetic Algorithm," in Proc. *Int. Conf. Genetic Algorithm*, pp. 416-421, 1991.
[27] W. Banzhaf, "Interactive Evolution," *Handbook of Evolutionary Computation*, 1997.
[28] J.-Y. Lee and S.-B. Cho, "Interactive Genetic Algorithm for Content-Based Image Retrieval," in *Proc. Asia Fuzzy Systems Symposium*, pp. 479-484, 1998.
[29] H. Takagi, "Interactive Evolutionary Computation: Fusion of the Capabilities of EC Optimization and Human Evaluation," Proc. *of the IEEE*, vol. 89, no. 9, pp. 1275-1296, 2001.

유 헌 우

1966년 12월 24일생. 1992년 인하대학교 전기공학과 졸업, 동대학 전기공학 석사(1994). 고려대학교 산업시스템정보공학 박사(2001). LG전자 생산기술센터기술개발 연구소(1994~1997). 코스모 정보통신 수석연구원(2000~2003). 현재 연세대학교 인지과학연구소 연구교수. 관심분야는 컴퓨터비전, 멀티미디어시스템, 제어이론

조 성 배

1988년 연세대학교 전산과학과(학사) 1990년 한국과학기술원 전산학과(석사) 1993년 한국과학기술원 전산학과(박사) 1993년~1995년 일본 ATR 인간정보통신연구소 객원연구원. 1998년 호주 Univ. of New South Wales 초청연구원. 1995년~현재 연세대학교 컴퓨터과학과 정교수. 관심분야는 신경망, 패턴인식, 지능정보처리