감정에 관련된 비디오 셧의 특징 표현 및 검출

강 행 봉† · 박 현 재††

요 약

인간과 컴퓨터간의 상호작용에 있어서 감정처리는 매우 중요한 부분이다. 특히, 비디오 정보처리에 있어서 사용자의 감정을 처리할 수 있다면 비디오 검색이나 요약본 추출 등 다양한 응용분야에 활용이 가능하다. 비디오 데이터로부터 이러한 감정 처리를 하기 위해서는 감정에 관련된 특징들을 표현하고, 검출하는 것이 필요하다. 쉽게 추출이 가능한 색상이나 모션 등의 저급 특징들로부터 고급 개념인 감정을 검출하는 것은 매우 어려운 일이지만, 감정에 관련된 여러 장면으로부터 LDA(Linear Discriminant Analysis)와 같은 통계적인 분석을 통해 감정에 관련된 특징들을 검출하는 것은 가능하다. 본 논문에서는 색상, 모션 및 셧 길이 정보로부터 감정과의 관련된 특징을 표현하고 검출하는 방법을 제안한다. 제안된 특징을 사용하여 감정 검출에 관련된 실험을 한 결과 바람직한 결과를 얻었다.

Representation and Detection of Video Shot's Features for Emotional Events

Hang-Bong Kang† · Hyunjae Park††

ABSTRACT

The processing of emotional information is very important in Human-Computer Interaction (HCI). In particular, it is very important in video information processing to deal with a user's affection. To handle emotional information, it is necessary to represent meaningful features and detect them efficiently. Even though it is not an easy task to detect emotional events from low level features such as colour and motion, it is possible to detect them if we use statistical analysis like Linear Discriminant Analysis (LDA). In this paper, we propose a representation scheme for emotion-related features and a detection method. We experiment with extracted features from video to detect emotional events and obtain desirable results.

키워드 : 감정 관련 특징 표현(Motion-Related Feature Representation), 감정 관련 특징 검출(Emotion-Related Feature Detection), LDA(Linear Discriminant Analysis)

1. 서 론

컴퓨팅 분야에 있어서 감정의 활용은 인간과 컴퓨터간의 상호 작용(HCI : Human Computer Interaction)에 있어 매우 중요한 역할을 한다 [1, 2]. 특히, 비디오 정보처리 분야에 있어서, 사용자 개인의 감정이나 취향이 반영된 한 차원 높은 효과적인 처리가 가능하다면, 다양한 응용이 가능해진다. 예를 들어, 내용 기반의 비디오 검색에 있어서 감정을 반영한 검색이 가능하다면 사용자가 제일 좋아하는 비디오 클립에 대한 검색, 슬픈 느낌을 주는 비디오 클립 등 감정이 결합된 검색 등이 가능해진다. 또, 사용자의 감정 상태에 적합한 비디오 데이터를 바탕으로 사용자 취향을 반영한 비디오 요약본도 만들 수 있다.

비디오 데이터에 관련된 사용자의 감정을 정확하게 처리하기 위해서는, 감정에 관련된 특징들을 추출 및 표현하여 적합한 계산 모델을 구축하는 것이 필요하다. 비디오 데이터로부터 추출이 가능한 저급 특징들로서는 일반적으로 색상, 모션, 질감, 형태 및 음성 정보 등이 있다. 이러한 저급 특징(low-level features)들로부터 고급 특징인 감정과의 관련성을 측정하여 적합한 특징을 찾아내는 것은 매우 어려운 일이지만, 감정에 관련된 여러 장면으로부터 통계적 분석을 통하여 이러한 저급 특징들과의 관계를 추론하는 것은 가능하다.

예를 들어, 색상 정보는 일반적으로 감정에 매우 밀접한 관계에 있다. 색상 자체가 고유의 감정을 나타내고 있고, 여러 개의 색상 조합 및 배열에 따라 다양한 감정을 발생시킨다[3-5]. 화가들이나 영화감독들은 자신의 감정을 색상

※ 본 논문은 정보통신부 정보통신연구진흥원에서 지원하고 있는 정보통신 기초기술 연구지원사업의 결과입니다.
† 정 회 원 : 가톨릭대학교 컴퓨터정보공학부 교수
†† 준 회 원 : 가톨릭대학교 대학원 컴퓨터공학과
논문접수 : 2003년 8월 23일, 심사완료 : 2004년 2월 6일

의 선택, 배열 및 조합을 통해 상대방에게 전달하고 있다[6, 7]. 모션 정보 역시 영화나 비디오의 시청자로 하여금 여러 가지 감정을 느끼게 하는데 커다란 역할을 하고 있다. 역동적인 움직임은 흥분된 감정을 표현하고 있고, 정적인 움직임은 차분한 감정을 느끼게 한다. 아울러, 반복되는 비디오 셧의 길이에 대한 패턴 즉 셧 컷율(shot cut rate)도 감정 표현에 적절하게 사용되고 있다. 따라서, 비디오 컨텐츠에 존재하는 색상, 모션 및 셧 컷율에 관한 정보의 효과적인 처리를 통해, 사용자가 느끼는 감정을 예측할 수 있다면, 감정 기반 비디오 정보처리가 가능해 진다.

따라서, 본 논문에서는 각 비디오 장면에서 느낄 수 있는 기본 감정을 검출하기 위해 비디오 데이터에 포함되어 있는 색상, 모션 및 셧 컷율에 대한 표현 및 감정에 관련된 요소의 검출에 관한 연구를 기술한다. 제2장에서는 감정 검출에 관련된 기존의 연구를 살펴보고, 제3장에서는 색상, 모션 및 셧 컷 율등의 저급 특징들에 대한 표현 방법에 대하여 기술하며, 제4장에서는 감정에 관련된 저급 특징들의 검출 방법에 대하여 기술한다. 제5장에서는 제안된 방법의 실험 결과에 대해 설명한다.

2. 감정 분석에 관련된 연구

최근들어 감정 분석을 응용한 연구가 지능적인 HCI를 위해 여러 연구기관에서 활발하게 진행되고 있으나, 아직까지 많은 어려운 문제들이 존재하고 있다. 감정이나 또는 감정의 지속적인 상태인 기분(mood)을 측정하기 위해서는 호르몬의 레벨, 신경 전달 속도 및 신경 시스템의 활동 상태 등을 측정하고, 이를 계량화하여 나타내는 것이 바람직하지만, 현재의 기술로는 구현하기 어렵다. 가능한 방법으로는 인간의 감정 시스템의 관찰로부터 얻은 감정이나 기분이 갖는 속성들을 바탕으로 측정하는 컴퓨팅 모델을 만드는 것이다[1]. 즉, 감정이 가지는 속성은 응답이 서서히 감소하고, 빠른 반복 속도를 갖는 연속적인 입력에 대해 감정의 강도는 증가하기 시작하며, 사람의 기질과 개인 특성에 따라 감정의 반응 정도가 달라진다. 또, 인간 감정 시스템은 비선형적이나 특정한 범위에서는 선형적으로 모델링 할 수 있고, 시불변이며, 충분한 입력이 있어야 반응하고, 어느 순간에 가서는 포화 상태에 이르게 된다. 더우기, 감정 시스템의 입력은 내부의 인지적인 또는 물리적인 프로세스에 의해 구동되고 이런 감정의 반응은 피드백을 제공하여서, 또 다른 감정이 발생하게 되며, 모든 입력은 배경이 되는 감정 상태에 영향을 미친다. 이런 상태의 생리학적인 신호를 측정하기 위해 Picard[1, 2]는 근전도(electromygram : EMG), 혈압(blood volume pressure), 피부전도도(galvanic skin response) 및 호흡 상태(respiration) 신호들을 사용하여 인간의 감정 상태를 측정하였다.

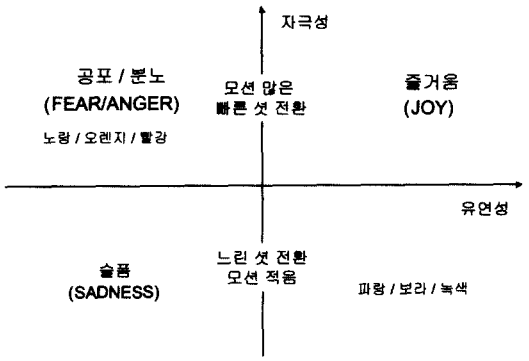
이러한 생리학적인 면 이외에 영화나 비디오 또는 그림을 감상하면서 느끼는 감정에 관한 연구도 많이 진행되었다. 색상 정보와 감정과의 관계는 화가나 영화를 제작하는 감독에 의하여 많이 연구되어 왔다. 그 중에서도 대표적인 것으로 Itten[4, 5]는 바흐하우스 운동에 화가로 참여한 경험을 바탕으로 표현 레벨에서 색상의 조합 결과를 표현하는 색상 언어에 관한 공식을 정의하였다. 그는 관찰자의 본능적인 관점에 관련 있는 이론을 개발하였다. Itten은 대비성, 일치성(accordance)를 이용하여 색상 정보를 표현하였다. 모션 및 사운드등 정보를 이용하여 감정 검출에 관한 연구도 진행되고 있다. Hanjalic and Xu[8]는 모션 및 사운드 정보를 이용하여 사용자의 감정 곡선을 구축하는 방법을 제안하였다. Moncrieff et al.[9]은 사운드에서 감정 분류 방법을 제안하였고, 이를 바탕으로 비디오 데이터로부터 공포 장면을 검출하였다.

저급 특징들로부터 고급 개념의 감정을 효과적으로 분류하는 것은 매우 어려운 작업이지만, 감정에 관련된 여러 장면으로부터 통계적인 분석을 통해 감정에 관련된 특징들을 검출하는 것은 가능하다. 본 논문에서는 비디오 데이터로부터 기본 감정을 분류하기 위한 저급 특징(색상, 모션, 셧 컷율)의 표현 및 검출하는 방법을 제안한다.

3. 감정에 관련된 저급 특징들의 표현

심리학에서는 감정을 일반적으로 공포(Fear), 분노(Anger), 슬픔(Sadness) 및 즐거움(Joy)등의 4개의 기본 감정으로 분류하고 있다[1]. 이러한 기본 감정을 표현하기 위해 사용되는 방법 중의 하나는 Lang[10]이 제안한 자극성(arousal)과 유발성(valence)의 2차원 좌표를 이용한 방식이다. 여기서 자극성은 감정의 흥분된 상태(excite)와 침착한 상태(calm)를 나타내고, 유발성은 긍정적인(positive) 상태와 부정적인(negative) 상태를 나타낸다. (그림 1)은 자극성과 유발성에 따라 분류한 감정의 예를 보여주고 있다[1, 2]. 행복이라는 감정은 흥분되면서도 긍정적이고, 분노나 두려움은 흥분되면서도 부정적인 면을 가지고 있다. 또, 색상 정보로 보면 노랑, 오렌지, 빨강색은 흥분과 부정적인 면에 가깝고, 파랑, 녹색, 보라의 색상 정보는 긍정적이며 침착한 면에 가깝다. 사진이나 영상으로 보면 골인 장면은 흥분하고 긍정적인 감정을 나타내고, 묘지 장면은 침착하면서 부정적인 면을 나타내는 경향이 있다. 공포와 분노의 감정인 경우 자극성과 유발성 관점에서 보면 비슷한 경향을 나타내므로, 색상 및 모션등의 저급 특징들로부터 이러한 감정의 분류는 매

우 어렵다. 따라서, 본 논문에서는 공포, 슬픔 및 즐거움 등의 세 개의 기본 감정의 검출만을 다룬다. 또, 이러한 감정들은 문화의 차이 또는 개인의 차이에 따라 다르게 느껴지기 때문에, 많은 사람들이 비디오나 영화를 보고 공감하는 감정과의 일반적인 특성을 찾는 방법을 찾는 것이 바람직하다.



(그림 1) 2차원 감정 스페이스

3.1 색상 정보 표현

일반적으로 빨강, 주황, 노랑 등의 파장이 긴 색들은 사람의 심장과 신경계의 움직임을 활발하게 동작하는 작용을 하며, 생리적 신경계가 활발해지면 이러한 색들을 보는 사람들의 상태 역시 활발해지며 기운이 넘치게 된다. 또한, 파란색에서 보라색으로 가는 파장이 짧은 색들은 심박수와 맥박을 낮추는 효과가 있어서 이러한 색들을 보는 사람들을 조용하고 기운이 없는 상태로 만들게 된다. 또, 각 색들은 고유의 심리적, 상징적 영향을 갖고 있다. 예를 들어, 검정색의 경우 공포, 분노, 죽음 등을 상징하게 되고, 회색은 우울, 무기력 등을 연상시킨다. 갈색의 경우는 대부분의 사람들이 자연의 색이라는 고정 관념을 갖고 있기 때문에 자연스럽고 편안한 분위기를 자아내는데 도움을 준다. 주황색의 경우는 발랄하고 활기 넘치는 기분을 느끼게 하고, 녹색은 안정되고 온화한 감정을 갖게 하는 특징을 갖고 있다[3].

색상 정보로부터 느끼는 감정을 표현하기 위해 Itten[5, 6]은 12개의 순수 색상을 동심원을 따라 배열하고 밝기 값의 변화를 위도에 따라 표현하였으며, 반지름이 증가하는 방향으로 채도를 증가하게 하였다. 반대되는 색은 원의 중심의 반대편에 배열하였다. 이러한 배열을 통해 대비성 및 조화(accordance)를 계산하였다. 본 논문에서는 Itten[5, 6]의 연구를 기반으로 11개의 문화권 색상(culture color)에 적용하였다. 문화권 색상이란 모든 문화권에서 색상을 기술하기 위해 공통적으로 사용되는 색상들을 뜻한다[11]. 모든 문화권에서 검정과 흰색에 상응하는 이름과 빨강, 노랑, 초록, 파랑에 대응되는 이름이 다른 어떤 것보다 먼저 사용되었다. 다른 색들에 대하여 사람들이 할당하는 이름들에 대한 분석을 통해 오직 11개의 색상(흰색, 검정색, 빨강, 초록, 노랑, 파랑, 갈색, 자주, 분홍, 오렌지색, 회색)에 일관되게 사용되어 이를 문화권 색상이라고 한다[11].

이러한 11개의 문화권 색상을 기본으로 색상 정보를 감정 레벨에서 표현하기 위해 다음과 같이 대비성(Contrast), 일치성(Accordance) 및 주된 색상(Dominant Color)등의 세 가지 특징들을 검출한다.

$$Color\_at\_emotion\text{-}level := \{Contrast,\ Accordance,\ Dominant\ Color\}, \quad (1)$$

$$where\ Contrast := \{Saturation,\ Light\text{-}dark,\ Warm\text{-}cold\}$$
$$Accordance := \{Complimentary,\ Harmony\}$$
$$Dominant\ Color := \{Single\ Dominant\ Color\}$$

대비성(Contrast)은 채도 대비성, 밝은 색과 어두운 색의 대비성 및 따뜻하고 차가운색의 대비성으로 표현한다. 일치성(Accordance)은 색상의 합이 하얀색이 되는 보색관계(complimentary) 및 색상들의 합이 회색이 되는 조화(Harmony) 관계로 표현할 수 있다. 또, 하나의 주된 색상도 감정을 표현하는데 중요하다. 이와 같은 특성을 계산하기 위해, 대표 프레임을 11개 문화권 색상으로 양자화하고, 임계 값보다 큰 연결된 영역을 5개 추출한다. 채도 대비성은 5개의 주된 영역의 채도를 구해 순서대로 나열한 다음, 채도가 큰 두 영역의 차이가 크지 않으면 합성하여 다음과 같이 합성된 영역의 채도를 구한다.

$$\mathrm{Merge}(A,B) = \frac{\mathrm{Sat.of.Region}_A \times \mathrm{Area.of.Region}_A + \mathrm{Area.of.Region}_A + \mathrm{Area.of.Region}_B}{\mathrm{Area.of.Region}_A + \mathrm{Area.of.Region}_B} + \frac{Sat.of.Region_B \times Area.of.Region_B}{Area.of.Region_A + Area.of.Region_B} \quad (2)$$

채도가 낮은 두 개의 주된 영역에서도 같은 작업을 반복한다. 채도가 높은 영역과 채도가 낮은 영역의 차이를 채도 대비성으로 정한다. 밝은 색과 어두운 색의 대비성도 이와 유사하게 5개의 주된 영역으로부터 구한다. 따뜻하고 차가운색의 대비성은 5개의 주된 영역에서 빨강, 핑크 및 오렌지색으로 구성된 따뜻한 색의 픽셀의 수와 초록, 파랑 및 보라로 구성된 차가운 색의 픽셀 수가 임계값 이상으로 존재할 때 대비성이 있다고 판단한다. 보색관계는 색들의 보색을 찾아서 판단하고, 조화 관계는 색상의 배열이 노랑-파랑-빨강이든지 녹색-오렌지-보라 등의 색상 판에서 배열 간격의 차이가 균등할 때 조화로운 배열로 판단한다[12]. 주된 색상은 배경이 한 개의 주된 색상으로 구성되어 있는지를 판단한다. Itten[5, 6]에 따르면, 따뜻한 색과 차가운색

의 대비성이나, 휴(Hue)의 대비성은 액션이나 역동적인 감정으로 표현되고, 초록색은 차분한 감정, 주된 색상의 존재나 대비성의 결핍은 불안함을 나타낸다고 하였다. 비디오 데이터로부터 감정관련 특징을 찾기 위해서는 식 (1)처럼 표현된 색상정보로부터 감정에 관련된 특징들을 찾아내는 것이 바람직하다.

3.2 모션 및 셧 컷 율의 정보 표현

Lang[10]의 2차원 감정 표현 중 자극성은 감정의 흥분된 상태(highly exciting) 및 침착한 상태(calm)를 나타내는데, 이러한 자극성은 모션 및 셧 컷 율를 가지고 표현할 수 있다. 먼저, 감정에 관련된 모션 정보로는 모션의 움직임으로 표현할 수 있고, 모션의 움직임은 모션의 크기 및 방향으로 정할 수 있다. 모션 정보는 다음과 같이 표현한다.

$$Motion\_at\_Emotion\_Level := \{Motion\_Activity\}, \quad (3)$$
$$where\ Motion\_Activity := \{Motion\_Intensity, Motion\_Phase\}$$

모션의 방향으로 본 논문에서는 "팬", "틸트", "줌" 및 "카메라모션 없음"의 네 가지 모션으로 표현한다.

셧 컷 율의 정보는 시청자에게 긴장감을 조절하는 요소로서 비디오 진행의 템포를 나타내고 있으며, 본 논문에서는 셧의 길이 정보로서 다음과 같이 표현한다.

$$Shot\_Cut\_rate := \{Shot\ Length\} \quad (4)$$

이러한 모션과 셧 컷율이 감정에 미치는 영향을 조사하기 위하여, 15개의 영화로부터 사용자가 느끼는 감정을 조사하였다. 감정을 분석한 결과는 <표 1>에 나타나 있으며 부분적으로 Picard[1]의 결과와 유사하다. 예를 들어, 공포의 감정인 경우, 줌이나 틸트인 경우가 많이 발생하였고, 모션 크기에는 큰 영향이 없으며 셧 컷 율은 매우 빠르다. 슬픈 감정인 경우 카메라 모션이 없는 경우가 많았으며, 셧 컷 율이 매우 느리다. 즐거운 감정인 경우 모션 크기가 매우 큰 특징이 나타난다. 식 (3)과 식 (4)로 표현된 모션과 셧 컷율의 특징중 감정에 관련된 특징을 검출하는 것이 중요하다.

〈표 1〉 모션 및 셧 컷율과 감정과의 관계

감 정	모 션 정 보	셧 컷율
공 포	줌, 틸트, 모션 크기는 상관없음	빠 름
슬 픔	카메라모션 없음 모션크기가 작음	느 림
즐거움	모션크기가 매우 큼	관련 없음

4. 감정에 관련된 저급 특징 정보의 검출

4.1 감정에 관련된 색상 정보 검출

색상의 특징들과 사람이 느끼는 감정과의 상관 관계를 계산하기 위해 먼저, 비디오 데이터를 색상 히스토그램과 모션 정보를 이용하여 비디오 셧으로 분할하고 비디오 셧을 대표하는 키 프레임을 찾는다[13]. 둘째로, 비디오 셧들을 공포, 슬픔 및 즐거움 등의 세 개의 기본 감정으로 분류한다. 비디오 셧에 느끼는 감정은 매우 주관적이므로, 본 논문에서는 10명이 느끼는 감정 중 7명 이상이 동일한 감정을 느낄 때 같은 감정이라고 분류한다. <표 2>는 15개의 비디오로부터 분류된 비디오 셧으로부터 분석한 세 개의 기본 감정과 색상과의 관계를 보여주고 있다. 예를 들어, 공포 감정을 갖는 비디오 셧의 경우 보통 어두운 푸른 색(dark and blue), 어두운 붉은 색(dark and red) 및 낮은 채도들의 색깔들이 많이 존재하고 있고, 즐거운 감정인 경우 밝은 색들로 이루어져 있으며, 슬픔의 경우 색상의 배열은 공포의 감정과 거의 유사하다. (그림 2)는 각 감정에 관련된 비디오 셧에 대한 대표 프레임을 보여 주고 있다. 실제로 식 (1)과 같은 색상 특징 중 분류하려고 하는 3개의 기본 감정에 관련된 특징이 있고, 별로 관련성이 없는 특징들도 있다. 따라서, 분류하려고 하는 감정과 관련된 특징들을 찾아낸 다음, 이를 바탕으로 감정 분류에 따른 특징들의 분포도를 작성하는 것이 바람직하다.

〈표 2〉 색상 정보와 감정과의 관계

감 정	색 상 정 보
공 포	어둡고 푸른색, 어둡고 붉은 색 낮은 채도의 색상
슬 픔	어둡고 회색빛깔의 색상 낮은 채도의 색상
즐거움	밝은 색상

비디오 셧으로부터 추출된 감정 관련 특징을 기반으로 세 개의 기본 감정을 분류하기 위해 본 논문에서는 Fisher의 Linear Discriminant Analysis(LDA) 방법을 사용한다[14, 15]. LDA는 특징 벡터들로부터 분류를 위한 가장 효과적인 주축을 찾는 것이다. d차원의 n개의 데이터 샘플 x_1, x_2, x_3, ⋯, x_n이 두 개의 클래스 데이터로 분리된다고 했을 경우, x_i의 컴포넌트에 대한 선형 결합으로서 다음과 같은 스칼라 내적을 얻는다.

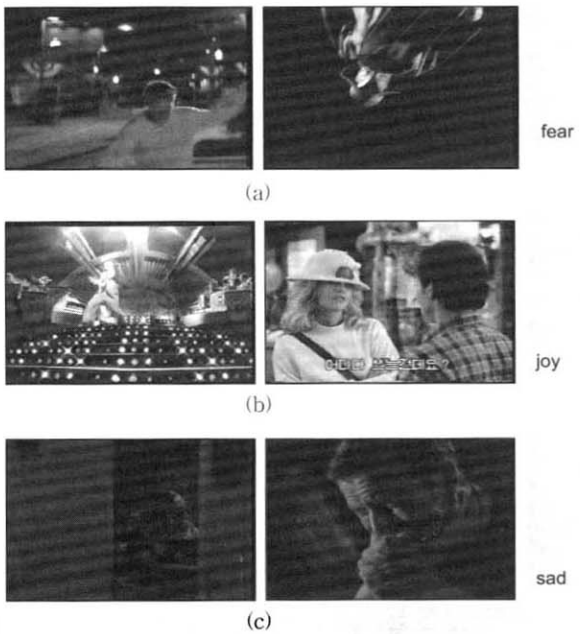
$$y = w^t x \quad (5)$$

이것에 대응하는 n개의 샘플 y_1, y_2, y_3, ⋯, y_n이 두 개의

집합으로 분류된다. 기하학적으로는 y의 방향을 가진 축에 x_i의 데이터가 투사된 것이 y_i 샘플이다. 여기서 y_1, y_2, y_3, …, y_n의 데이터를 두 개의 클래스로 잘 분리할 수 있는 w의 방향을 찾는 것이 중요하다. (그림 3)은 두 개의 클래스의 분류가 용이한 축을 보여주고 있다. 이런 방향을 찾기 위해 S_W(within-class matrix)와 S_B(between-class matrix)를 다음과 같이 계산한다.

$$S_W = \sum_{i=1}^{2} S_i \text{ where } S_i = \sum_{X \in D_i} (x - m_i)(x - m_i)^t \quad (6)$$

$$S_B = \sum_{i=1}^{2} n_i (m_1 - m_2)(m_1 - m_2)^t \text{ where } m_i = \frac{1}{n_i} \sum_{X \in D_i} X \quad (7)$$



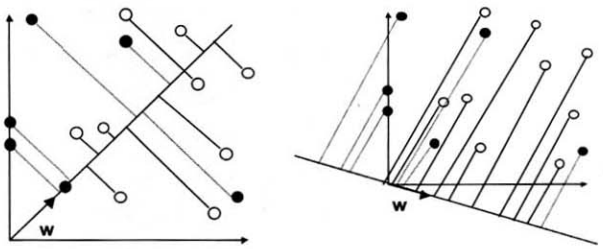
(그림 2) 각 감정에 관련된 비디오 셧의 예 :
(a) 공포, (b) 즐거움, (c) 슬픔

$$J(w) = \frac{w' S_B w}{w' S_w w} \quad (8)$$

Fisher의 LDA 방법은 대부분의 데이터들에서 정확한 선형 분리 결과를 보인다. 그러나 두 클래스의 데이터 크기의 차이가 클 경우, 오분류가 되는 결과가 가끔 발생한다. 이를 해결하기 위해 클래스 내의 분산 정보를 구할 때, 단순히 두 클래스의 분산의 합을 이용하지 않고, 전체 데이터 개수에 대한, 각 클래스의 데이터 개수의 비를 가중치로 적용하여, 클래스의 크기 정보가 고려될 경우, 정확한 결과를 얻을 수 있다. 즉, 식 (6)을 다음과 같이 변형할 수 있다.

$$S_w = \frac{1}{N_1} S_1 + \frac{1}{N_2} S_2 \quad (9)$$

LDA를 이용한 선형 분류 방법은 일반적으로 고차원 공간에 존재하는 데이터를 저차원 공간에서 클래스 별로 분류할 수 있도록 변환할 수 있는 장점을 가진다. 이러한 이유로 선형적인 분류에서 널리 사용되고 있으나, 몇 가지 문제점을 가진다. 먼저 학습 데이터와 테스트 데이터가 상이할 경우 학습 데이터에 의해 생성된 저차원의 공간상에서 테스트 데이터가 수렴하지 않는다. 또한, 학습 영상의 수가 적을 경우 고유값(eigenvalue)과 고유 벡터(eigenvector)를 계산하는 알고리즘이 수렴하지 않는다. 이러한 단점을 해결하기 위하여 PCA(Principal Component Analysis)와 LDA를 결합한 형태의 기법을 도입한다.



(그림 3) LDA의 예 (a) 분리가 잘되지 않음,
(b) 비교적 분리가 잘됨.

이 기법은 PCA를 이용하여 먼저 원 데이터를 저차원으로 투영하고, 그 결과를 이용하여 LDA를 수행한다. 이를 수식으로 나타내면 다음과 같다.

$$y_i = W_{LDA}^t W_{PCA}^t x_i \quad (10)$$

W_{LDA}는 LDA에 의해 생성되는 공간의 생성벡터(Basis) 집합이며 W_{PCA}는 PCA에 의해 생성되는 공간의 생성 벡터(basis)이다. x_i는 입력 벡터이며 y_i는 PCA와 LDA를 거친 후의 결과 벡터이다. 본 논문에서는 감정 분류에 있어서 LDA 방식과 PCA+LDA 방식을 사용한다.

또, 지금처럼 3개의 감정과 이들을 제외한 나머지 감정으로 이루어진 4개의 클래스를 분류하기 위해서는 LDA 방식을 일반화한 Multiple Discriminant Analysis(MDA) 방식이 필요하다[15]. c개의 클래스를 구분하는 MDA에서도 S_W(within-class matrix)와 S_B(between-class matrix)를 다음과 같이 일반화할 수 있다. 식 (2)는 다음과 같이 d-by-(c-1)의 행렬 W로 할 수 있고, 프로젝션시켰을 경우 다음과 같이 된다.

$$y = W^t x \quad (11)$$

따라서, 평가 함수는 다음과 같이 되고, 궁극적으로 J(·)

를 최대로 하는 W를 구하면 된다.

$$J(W) = \frac{|S_B'|}{|S_W'|} = \frac{|W^t S_B W|}{|W^t S_W W|} \quad (12)$$

4.2 감정에 관련된 모션 및 셧 컷율 검출

모션 움직임(Moiont_Activity)은 모션의 크기(Motion_Intensity) 및 모션의 방향(Motion_Phase)으로 표현할 수 있고, 셧 컷 율은 셧 길이의 시간 축 상의 패턴으로 표현할 수 있다. 모션의 크기 및 방향을 구하기 위해서는 연속된 두 프레임 사이의 프레임 차이(frame difference)를 계산하고, 이 프레임 차이가 임계 값을 넘는 경우 프레임을 9개의 블록으로 분할한다. 분할된 블록으로부터 Lucas-Kanade[16] 방식의 옵티칼 플로우를 계산하여 8개의 방향으로 양자화한 다음 각 블록의 모션의 방향을 정한다. 모션 크기는 주된 움직임 벡터로부터 계산한다. 만약 한 블록 내에서 방향 θ를 갖는 옵티칼플로우 벡터(m(x, y))의 개수(N_k)가 블럭 전체 옵티칼플로우 벡터의 개수의 합(N_{total})의 반 이상이 될 경우 이 방향을 영역의 대표 모션 벡터의 방향(Motion Phase)으로 간주한다. 또, 모션 벡터의 크기(Motion Intensity)는 대표 모션 벡터와 같은 방향의 옵티칼플로우 벡터의 크기를 평균하여 계산한다. 즉,

$$\text{Motion Phase} = \begin{cases} \theta & \text{if } N_k > N_{total}/2 \\ don't\ care & \text{otherwise} \end{cases} \quad (13)$$

$$\text{Motion Intensity} = \sum_{i=1}^{Nk} m(x,y)/N_k \quad (14)$$

이렇게 하여 프레임의 각각의 블럭에 존재하는 대표 모션 벡터의 방향과 모션 크기를 계산한다. 각 비디오 프레임의 움직임은 블럭의 모션 방향을 이용하여 템플레이트 매칭을 통해 "팬(pan)", "줌(zoom)", "틸트(tilt)" 및 "카메라모션 없음(no camera motion)" 등으로 구분한다. "카메라 모션 없음" 경우에는 오브젝트의 모션 크기를 프레임 차이로 계산한다. 비디오 셧의 모션은 각 프레임의 모션을 구한 후 가장 많은 수의 프레임이 갖는 방향으로 정한다.

셧 컷 율은 시간축 상의 셧의 길이 패턴으로 표현하는데 각 셧의 길이는 전체 비디오 셧을 길이를 기준으로 분류하여 메디안(median)을 구한 다음 이를 기준으로 정규화한다. 왜냐하면, 메디안이 평균 셧의 길이보다 잡음의 영향을 덜 받는 결과를 보여줄 수 있기 때문이다. 셧의 길이 x'를 구하는 식은 다음과 같다.

$$x_i' = \frac{x_i - A}{B - A} \quad (15)$$

여기서 x_i는 주어진 셧의 길이이고, A 및 B는 메디안으로부터 일정 범위에 존재하는 임의로 정한 두 점이다.

여기서 구한 모션 및 셧 컷율 정보도 4.1절에서 기술한 바와 같이 PCA+LDA 방식 및 MDA 방식을 적용하여 감정에 관련된 정보를 검출한다.

5. 실험 결과

5.1 데이터 분석

색상, 모션, 셧 컷 율 등의 저급 정보를 사용하여 감정을 분석하기 위하여, 공포, 슬픔 및 즐거움이 포함된 15편의 영화를 선택하였다. <표 3>은 실험을 위해 사용된 영화 목록이다. 이 영화들을 대표 프레임-비디오 셧-비디오 장면(scene)으로 분할하여, 각각의 영화에 대하여 감정을 분석하였다[13, 17]. 사람마다 느끼는 감정은 주관적이어서, 10 사람

〈표 3〉 테스트 영화 목록

번호	영화 제목	시 간	감정 분류		
			공포	슬픔	즐거움
1	(다잉 영 하편) Dying Young -part 2	46 : 57	0	26	23
2	(링 상편) Ring -part 1	50 : 20	56	0	0
3	(링 하편) Ring -part 2	41 : 50	30	0	0
4	(뉴욕의 가을) Autumn In New York	26 : 34	74	0	0
5	(나 홀로 집에) Home Alone	61 : 32	0	0	62
6	(나는 네가 지난 여름에 한 일을 알고 있다 상편) I know what you did in last summer -part 1	54 : 53	95	0	0
7	(나는 네가 지난 여름에 한 일을 알고 있다 하편) I know what you did in last summer -part 2	40 : 29	392	0	0
8	(쥬라기 공원 상편) Jurasic Park -part 1	58 : 33	0	0	15
9	(쥬라기 공원 하편) Jurasic Park -part 2	60 : 46	152	0	6
10	(마스크 상편) Mask -part 1	57 : 04	0	0	94
11	(마스크 하편)Mask -part 2	36 : 40	0	0	96
12	(라이언 일병 구하기) Saving Private Ryan	41 : 07	159	0	0
13	(스크림) Scream	57 : 47	179	0	0
14	(타이타닉) Titanic	61 : 36	0	31	0
15	(해리가 샐리를 만날 때) When Harry met Sally	41 : 30	0	31	0
	TOTAL		1137	88	296

이 감상하고, 이중 적어도 7명 이상이 공통으로 느끼는 장면을 감정으로 분류하였다. 실험에 참가한 10명의 구성은 20대 6명, 30대 3명, 40대 1명이며, 남자가 9명 여자가 1명이다. 실험에 사용된 15개의 영화는 다양한 테스트를 위해 4개의 집합(C, D, E, F)으로 구성하였다. 공포셧의 개수가 슬픔이나 즐거움 보다 많으나 각 감정에 따라 테스트하므로 전체적인 감정 분류에는 별다른 영향을 주지 않는다고 생각된다. 4개의 집합은 각 감정(공포, 슬픔, 즐거움)이 균형적으로 배분하도록 구성하였으며, <표 4>는 각 집합이 갖고 있는 감정별 비디오 셧의 개수를 보여주고 있다. <표 5>는 추출된 세 가지의 감정을 갖고 있는 장면의 예를 보여주고 있다. 각 데이터 집합으로부터 식 (1), 식 (3) 및 식 (4)로 표현된 색상, 모션 및 셧 컷 율의 감정에 대한 중요도를 실험하였다. 각 집합에 대하여 기본 감정과 각 특징들의 상관 관계를 실험한 결과, 색상 정보의 경우 채도 대비성, 밝은색과 어두운 색의 대비성 및 주된 색상 특징들이 감정을 결정하는데 중요한 역할을 하는 것을 알 수 있었다. 모션 정보의 경우 모든 특징들이 감정에 대하여 유사한 상관관계를 보여주었고, 셧 컷 율 역시 감정에 관련성이 매우 높았다. 이러한 특징들이 감정 분류에 사용될 수 있는 가능성을 증명하기 위해 LDA, PCA 및 MDA를 이용하여 검증하였다.

〈표 4〉 테스트 데이터 집합

Title	C			D			E			F		
	F	S	J	F	S	J	F	S	J	F	S	J
뉴욕의 가을		40			34			74				
나 홀로 집에			30			32						62
나는 네가 지난 여름에 한 일을 알고 있다(상)	45			50			95					
나는 네가 지난 여름에 한 일을 알고 있다(하)	192			200			392					
쥬라기 공원(상)			10			5			15			15
쥬라기 공원(하)	80		3	72		3			6	152		6
마스크(상)			50			44						94
마스크(하)			50	26		46			96			
링(상)	30			20			56					
링(하)	10			79						30		
라이언 일병 구하기	80			90			169					
스크림	89									179		
타이타닉		15			16		31					
해리가 샐리를 만날 때			16			108						124
다잉 영(하)		13			13		26					
감정 셧수	526	68	159	537	63	238	769	74	107	361	25	301

〈표 5〉 감정 관련 장면의 예

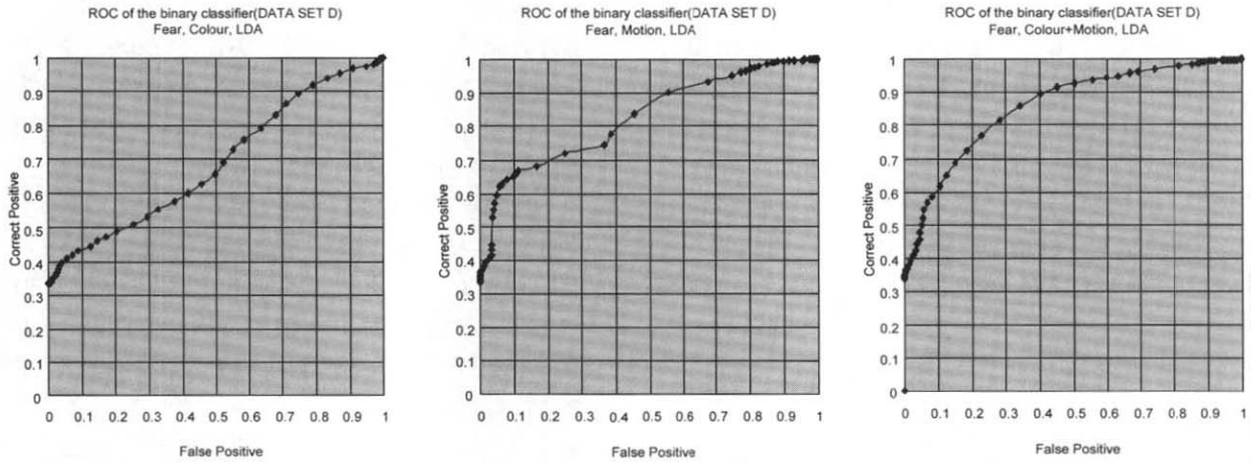
영 화	감 정	특 징	설 명
쥬라기 공원 (하)	공 포	어두운 배경, 카메라 상하 움직임	정전으로 인해 전기자동차가 '티라노사우르스'가 있는 구역안에서 정지한 후 티라노사우르스가 일행들을 공격함
다잉 영	슬 픔	어두운 배경, 줌	Hillary와 Victor가 대화를 하고 있음
해리가 샐리를 만났을때	즐거움	밝고 긴 셧, 카메라 움직임 없음	Harry와 Sally가 사랑에 대하여 이야기 하고 있음.
쥬라기 공원(하)	공 포	어두운 배경, 빠른 움직임, 셧이 빠르게 바뀜	박사 일행들을 구하기 위해 갔던 사람들도 '티라노사우르스'의 공격을 받음

5.2 LDA 및 MDA를 적용한 감정 분류

색상 정보로는 채도 대비성, 밝은 색-어두운 색 대비성 및 주된 색상을 검출하였고, 모션은 "팬", "틸트", "줌" 및 "카메라 모션 없음"의 방향 및 크기를 검출하였다. 셧 길이 정보는 메디안을 기준으로 계산하였다. 추출된 정보를 이용하여 LDA를 이용하여 분석하였다. LDA를 사용하여 각각의 감정 검출율에 대해 ROC(Receiver Operating Characteristic) 곡선을 작성하였다. 색상이나 모션 특징만 사용하였을 경우, 색상+모션 정보를 사용하였을 경우에 따라 ROC 곡선을 계산하였다. (그림 4), (그림 5), (그림 6)은 각각의 감정에 대한 ROC 곡선이다. False positive가 30% 미만 일 때 70% 이상의 검출 율을 얻는 것을 알 수 있다. 색상이나 모션특징을 단독으로 하였을 경우보다 색상+모션특징을 사용하였을 경우 좀더 좋은 결과를 얻었다. (그림 7)은 LDA+PCA의 경우의 결과를 보여주고 있다. 일반적으로 LDA를 사용하였을 경우보다 약간 더 나은 결과를 얻었다. 또, (그림 8)은 MDA를 사용한 분포도로서 본 논문에서 제안한 특징 표현을 이용하여 감정 검출이 가능하다는 것을 보여주고 있다.

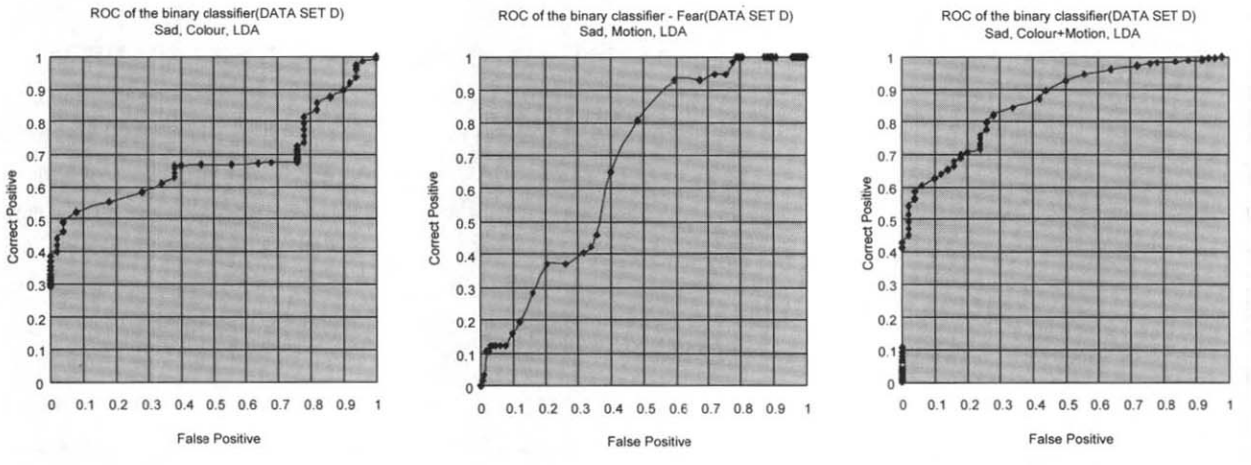
6. 결 론

본 장에서는 비디오로부터 색상, 모션 및 셧 컷율 정보의 저급 특징을 이용하여 공포, 슬픔 및 즐거움 등의 3가지 기본 감정을 검출하기 위한, 특징들의 표현 및 검출 방법을 제안하였다. 색상의 특징으로는 대비성, 일치성(accordance) 및 주된 색상을 사용하여 표현하였고, 모션 특징으로는 모션의 방향과 크기 및 셧의 시간 축 상의 길이 패턴을 사용하였다. 이러한 특징의 표현으로부터 감정에 관련된 특징을 추출한 후 LDA 방식을 적용하여 감정을 검출하였다. 3개



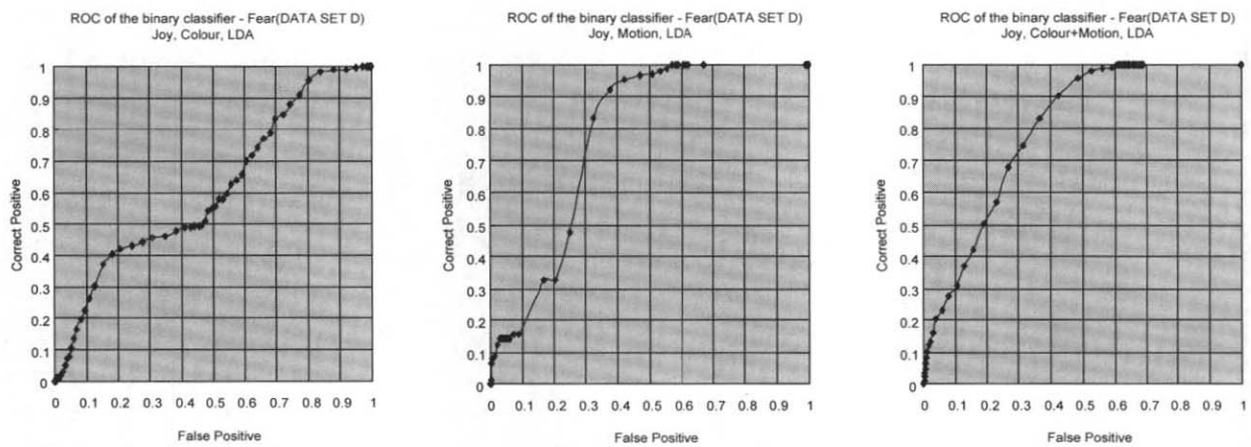
(a) 색상 (b) 모션 (c) 색상 + 모션

(그림 4) LDA를 이용한 공포 감정의 ROC 곡선



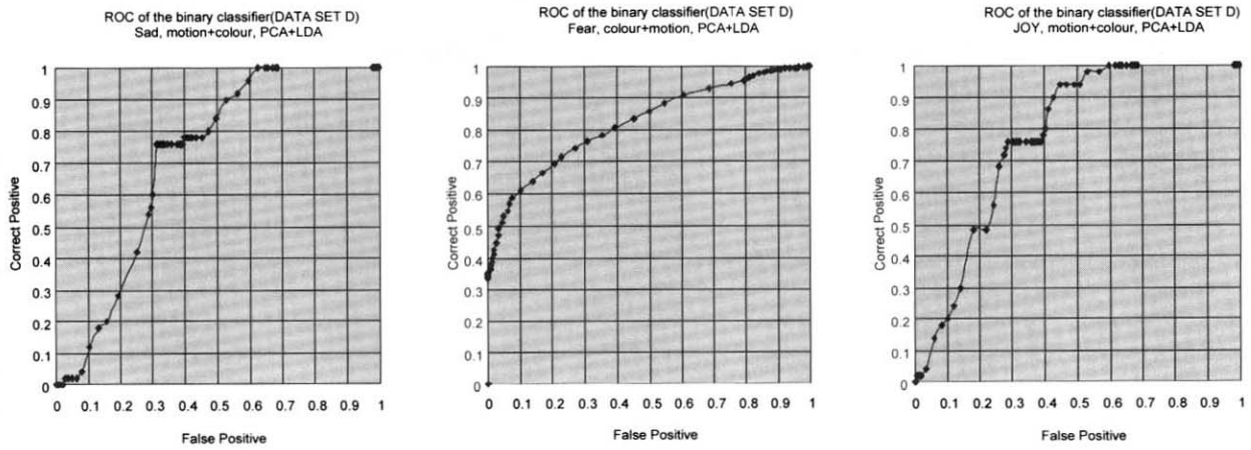
(a) 색상 (b) 모션 (c) 색상 + 모션

(그림 5) LDA를 이용한 슬픔 감정의 ROC 곡선



(a) 색 상 (b) 모 션 (c) 색상 + 모션

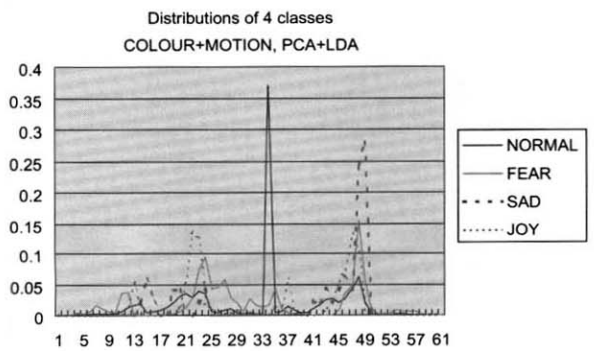
(그림 6) LDA를 이용한 즐거움 감정의 ROC 곡선



(a) 공포(색상 + 모션) (b) 슬픔(색상 + 모션) (c) 즐거움(색상 + 모션)

(그림 7) PCA + LDA를 이용한 각 감정분류에 대한 ROC 곡선

의 기본 감정을 표현한 15편의 영화에 적용한 결과, 30%미만의 false positive인 경우 75%이상의 검출 율을 갖는 바람직한 결과를 얻었다. 색상이나 모션만의 특징보다 색상 + 모션의 특징을 사용하였을 경우 보다 더 바람직한 결과를 얻었다. 또, PCA + LDA 방식이 LDA 방식보다 일반적으로 더 나은 결과를 얻을 수 있었다. 또, MDA를 사용하였을 경우 4개의 감정의 분리가 가능하다는 것을 알 수 있었다. 이러한 감정 분류를 보다 더 정확히 하게 하기 위해서는 오디오 정보 및 텍스트 정보(closed caption)를 덧붙여 멀티 모달 방식의 시스템을 구현하는 것이 바람직하다.



(그림 8) MDA를 이용한 감정별 분류 결과

참 고 문 헌

[1] R. Picard, *Affective Computing*, MIT Press, 1997.

[2] R. Picard, "Affective Computing for HCI," *Proc. of HCI'99*, Aug., 1999.

[3] M. Cooper, *Color Smart : How to use color to enhance your business and personal life*, Pocket Book, 2000.

[4] J. Itten, *The Art of Color*, Wiley, 1973.

[5] J. Itten, *The Elements of Color*, Wiley, 1970.

[6] J. Monaco, *How to read a film*, Oxford, 2000.

[7] C. Dorai and S. Venkatesh eds. *Media Computing : Computational Media Aesthetics*, Kluwer Academic Publishers, 2002.

[8] A. Hanjalic and L. Xu, "User-oriented Affective Video Content Analysis," *Proc. IEEE Workshop on Content-Based Access of Image and Video Library*, Kauai, HI, pp.50-57, Dec., 2001.

[9] S. Moncrieff, C. Dorai, and S. Venkatesh, "Affect Computing in Film through Sound Energy Dynamics," *Proc. ACM Multimediap'01*, pp.525-527, 2001.

[10] P. Lang, "The emotion probe : Studies of motivation and attention," *American Psychologist*, 50(5), PP.372-385, 1995.

[11] E. Goldstein, *Sensation and perception*, Brooks/Cole, 1999.

[12] A. D. Bimbo, *Visual Information Retrieval*, Morgan Kaufmann, 1999.

[13] H. Zhang, J. Wu, D. Zhong and S. Smoliar, "An Integrated System for Content-based Video Retrieval and Browsing," *Pattern Recognition*, 30(4), 1997.

[14] R. Duda, P. Hart and D. Stork, *Pattern Classification*, Wiley-Interscience, 2002.

[15] S. Gong et al, *Dynamic Vision*, Imperial College Press, 2000.

[16] B. Lucas and T. Kanade, "An Iterative Technique of Image Registration and Its Application to Stereo," *Proc. IJACI*, pp.674-679, 1981.

[17] H. Sundaram and S. Chang, "Determining Computable Scenes in Films and their Structures using Audio-Visual Memory Models," *Proc. Acm Multimedia'00*, 2000.

강 행 봉

e-mail : hbkang@catholic.ac.kr

1980년 한양대학교 전자공학과졸업

1986년 한양대학교 대학원 전자공학과 (석사)

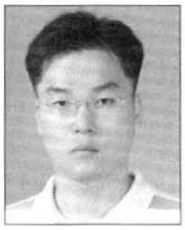
1989년 미국 Ohio State University 컴퓨터 공학 석사

1993년 미국 Rensselaer Polytechnic Institute 컴퓨터 공학 박사

1994년~1997년 삼성 종합기술원 수석 연구원

1997년~현재 가톨릭대학교 컴퓨터정보공학부 부교수

관심분야 : 컴퓨터 비젼, 멀티미디어 시스템, 인공 지능, 생체 인식 및 Bioinformatics



박 현 재

e-mail : hyunjapark@catholic.ac.kr

2002년 가톨릭 대학교 컴퓨터공학과

2002년~현재 가톨릭대학교 컴퓨터공학과 석사과정

관심분야 : 패턴 인식, 기계 학습, 모델 기반 영상 처리