

# 칼라 정보를 이용한 비디오 데이터에서의 감정 분석

전성호, 강행봉  
가톨릭대학교 컴퓨터공학과  
e-mail: sungho35@catholic.ac.kr

## Emotional analysis in video data using color information

Sung-Ho Chun, Hang-Bong Kang  
Dept. of Computer Engineering, The Catholic University of  
Korea

### 요 약

사용자의 감정에 관련된 정보를 처리하는 것은 인간과 컴퓨터와의 상호작용(HCI)에 있어서 매우 중요한 역할을 한다. 특히 비디오 데이터에 대한 사용자의 감정을 처리하는 것은 비디오 검색이나 요약 본 구성에 매우 중요하다. 사용자의 감정을 처리하기 위해서는 감정에 관련된 특징들을 추출 및 측정하고 이를 기반으로 비디오 장면을 분류하는 것이 필요하다. 본 논문에서는 칼라 정보를 바탕으로 Fisher의 Linear Discriminant Analysis 방식 및 Mahalanobis Distance 측정을 이용하여 기본 감정의 분류 방식을 제안한다. 공포 감정의 경우 77.8%의 의미 있는 결과를 얻었다.

### 1. 서론

컴퓨팅 분야에 있어서 감정의 활용은 인간과 컴퓨터간의 상호 작용(HCI: Human Computer Interaction)에 있어 상당한 영향을 미칠 것이다 [1-3]. 특히, 비디오 정보처리 분야에 있어서, 사용자 개인의 감정이나 취향이 반영된 한 차원 높은 효과적인 처리가 가능하다면, 다양한 응용이 가능해진다. 예를 들어, 사용자가 제일 좋아하는 비디오 클립에 대한 검색, 슬픈 느낌을 주는 비디오 클립 등 감정이 결합된 검색등이 가능해지고, 사용자의 감정 상태에 적합한 비디오 요약본도 만들 수 있다.

비디오 데이터에 관련된 사용자의 감정을 정확하게 처리하기 위해서는, 감정에 관련 된 특징들을 추출 및 측정하는 기술이 필요하다. 비디오 데이터로부터 추출이 가능한 저급 특징들로서는 칼라 정보나 모션 정보 및 음성 정보 등이 있다. 이러한 저급 특징(low-level features)들로부터 고급 특징인 감정을 분류한다는 것은 매우 어려운 일이지만, 각각의 특징들이 감정에 미치는 영향을 바탕으로 매핑하는 것

이 가능하다. Hanjalic 및 Xu [4]는 모션 및 사운드 정보를 이용하여 감정 곡선을 추출하여, 사용자의 감정을 분석하였다. Moncrieff et al. [5]은 사운드 에너지로부터 감정에 관련된 사운드를 분류하는 방법을 제안하여 공포영화에 있어서 공포 감정을 갖는 장면을 분류하였다.

칼라 정보 역시 감정에 매우 밀접한 관계에 있다. 칼라 자체가 고유의 감정을 나타내고 있고, 또 칼라의 조합 및 배열에 따라 다른 감정을 발생 시킨다 [6]. 즉, 화가들이나 영화감독들은 자신의 감정을 칼라의 선택, 배열 및 조합을 통해 상대방에게 전하고 있다.

따라서, 비디오 콘텐츠에 존재하는 칼라 정보의 효과적인 처리를 통해 사용자가 느끼는 감정을 예측할 수 있다면, 감정 기반 비디오 정보처리가 가능해진다.

따라서, 본 논문에서는 비디오 데이터에 포함되어 있는 칼라 정보를 추출 및 분석하여 각 비디오 셋에서 느낄 수 있는 기본 감정을 검출하는 방법을 제안

한다. 제 2절에서는 칼라 정보와 감정과의 관계를 기술하며 제 3절에서는 감정에 관련된 칼라정보의 추출에 대하여 Fisher의 LDA (Linear Discriminant Analysis) 방식과 Mahalanobis Distance에 의한 방식에 대해 설명하며 제 4절에서는 제안된 방법의 실험 결과에 대해 기술한다.

**2. 칼라와 감정과의 관계**

일반적으로 빨강, 주황, 노랑 등의 파장이 긴 색들은 사람의 심장과 신경계의 움직임을 활발하게 하는 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 있고, 그로 인해서 이러한 색들을 보는 사람들을 조용하고 기운이 없는 상태로 만들게 된다.

또한 각 색들은 고유의 심리적, 상징적 영향을 갖고 있는데 예를 들어 Black의 경우 공포, 분노, 죽음 등을 상징하게 되고, Gray는 우울, 무기력 등을 연상시킨다. Brown의 경우는 대부분의 사람들이 자연의 색이라는 고정관념을 갖고 있기 때문에 자연스럽고 편안한 분위기를 자아내는데 도움을 주며, 주황색의 경우는 발랄하고 활기 넘치는 기분을 느끼게 하고, 녹색은 안정되고 온화한 감정을 갖게 하는 특징을 갖고 있다.

비디오 데이터로부터 발생하는 감정을 표현하기 위해 많이 사용되는 방식은 arousal과 valence의 2차원 좌표이다 [7]. 여기서 arousal은 감정의 흥분된 상태 (excite)와 침착한 상태 (calm)를 나타내고, valence는 긍정적인(positive) 상태와 부정적인(negative) 상태를 나타낸다.

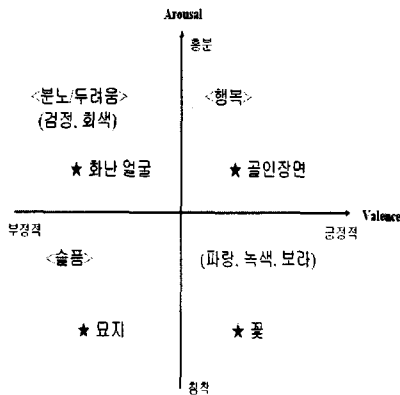


그림 1: arousal과 valence의 2차원 좌표

그림 1은 arousal과 valence에 따라 분류한 감정, 칼라 및 영상의 예를 보여주고 있다 [1,3]. 행복이라는 감정은 흥분되면서도 긍정적이고, 분노나 두려움은 흥분되면서도 부정적인 면을 가지고 있다. 또, 칼라 정보로 보면 노랑, 오렌지, 빨강색은 흥분과 부정적인 면에 가깝고, 파랑, 녹색, 보라의 칼라 정보는 긍정적이면 침착한 면에 가깝다. 사진이나 영상으로 보면 골인 장면은 흥분하고 긍정적인 감정을 나타내고, 묘지 장면은 침착하면서 부정적인 면을 나타내는 경향이 있다.

**3. 칼라 정보의 추출**

일반적으로 기본 감정으로는 공포, 분노, 즐거움 및 슬픔 등을 들 수 있는데 공포 및 분노는 그림 1에서처럼 감정 스페이스에서 같은 영역에 존재하므로 구분하기가 매우 까다롭다. 본 논문에서는 공포, 즐거움 및 슬픔의 감정에 관련된 칼라 정보를 추출한다.

먼저, 인간의 지각과 유사한 HSV 칼라 스페이스를 사용한다. 색상과 감정과의 관계를 추출하기 위해 H(Hue) 정보는 11개의 culture color(white, black, gray, red, yellow, green, blue, brown, purple, pink, orange)로 양자화 하였다 [8]. 여기에 채도 (Saturation) 및 밝기(Value)를 포함하여, 비디오 데이터를 분석하였다. 비디오 학습 데이터로 6편의 영화를 선택하여 공포, 슬픔, 즐거움의 감정을 갖는 비디오 장면을 분류하였다. 이때 사람마다 느끼는 감정이 일치하지 않으므로 분류한 5명중 3명이상이 같은 감정을 갖는 장면을 감정 장면으로 분류하였다.

학습 데이터의 분석을 통해 공포, 슬픔 및 즐거움 감정은 각각의 culture color 보다는 black과 gray로 구성된 어두운 칼라, yellow, green, brown, pink, orange로 구성된 밝은 칼라, Saturation 및 Value가 주된 요소라는 것을 알았다. 즉, black 및 gray로 구성된 칼라는 공포, 우울, 분노 등의 감정에 관련이 깊고, yellow, green, brown, pink 및 orange로 구성된 칼라의 경우는 자연스러움, 활발함, 안정 등의 감정과 관련이 깊다.



그림 2: 어두운 칼라 및 밝은 칼라의 예

감정에 관련된 칼라 특징을 기반으로 비디오 데이터의 감정을 분류하기 위해서 Fisher의 LDA 방식 및 Euclidean Distance 대신 Mahalanobis Distance를 이용한 감정 분류 방식을 적용 한다 [9].

LDA에서는 선형적 분류가 가장 용이한 주축을 찾은 다음 그 주축에 입력된 자료를 projection 시켜 주축 상에서 자료가 어떤 경향을 갖고 분포되어 있는지를 분석한다. 이 때 찾은 주축은 각 feature 축과 이루는 각도에 의해서 나타나고, projection은 다음과 같은 방법에 의하여 실행된다. 식 (1)의  $W_p$ 는 projection 된 자료의 feature vector이고,  $S$ 는 입력된 자료의 feature vector이며,  $W$ 는 projection 될 축의 feature vector이다. 식 (1)에 의해 구해진  $W_p$ 는  $\|W_p\|$ 에 의해서 projection 된 주축의 좌표로 매핑 된다.

$$W_p = \frac{S \cdot W}{\|W\|^2} \cdot W \quad (1)$$

또 다른 감정 분류 방법에서는 각 감정별로 분류된 데이터를 클러스터링 한 후 입력된 자료와의 Mahalanobis Distance 방식을 적용하여, 입력된 자료와 가장 가까운 거리에 위치한 클러스터로 분류한다. Mahalanobis Distance는 각 feature의 평균과 공분산 행렬을 이용하여 입력된 자료와 각 클러스터 사이의 거리를 측정하는 방식으로 식 (2)와 같은 방법에 의하여 구할 수 있다. 식 (2)에서  $x$ 는 feature vector이고,  $m_x$ 는  $x$ 의 평균이며,  $C_x$ 는  $x$ 의 공분산 행렬이다.

$$r^2 = (x - m_x)' C_x^{-1} (x - m_x) \quad (2)$$

#### 4. 실험결과

실험 비디오 데이터는 I know what you did last summer의 전반부(summer1) 및 후반부(summer2), Dying Young 후반부(dying), Autumn in New York 후반부(autumn), When Harry met Sally 후반부(harry), Scream 전반부(scream) 그리고 Mask 전반부(mask)를 대상으로 생성되었으며, 각 영화에 대한 정보는 다음의 표 1과 같다.

실험을 위해서 비디오 데이터로부터 각 장면들을 분리해 내고, 장면들을 구성하는 컷들을 구분한 후 한 컷이 포함하고 있는 여러 장의 프레임들 중 하나를 추출하여 그 프레임을 컷을 대표하는 대표 프레임으로 정한다. 그러나 일반적으로 전경보다는 배경이 감정에 더

큰 영향을 미치는 것으로 알려져 있으므로 추출된 대표 프레임을 배경과 전경으로 분리하여 배경의 가중치를 전경의 가중치 보다 더 높게 조정하였다. 이렇게 가중치가 부여된 대표 프레임을 11가지의 Hue로 양자화하여 사용하고, 그림 3은 영화 속의 한 대표 프레임을 11가지의 Hue로 양자화한 예를 보여주고 있다.

표 1: 각 영화별 정보

영화제목	감정	시간	컷 수	감정 컷 수
summer1	공포	55:01	645	273
summer2	공포	43:41	821	587
dying	슬픔	48:04	440	56
autumn	슬픔	26:33	242	74
harry	즐거움	43:34	333	124
scream	공포	57:47	721	381
mask	즐거움	57:04	697	224

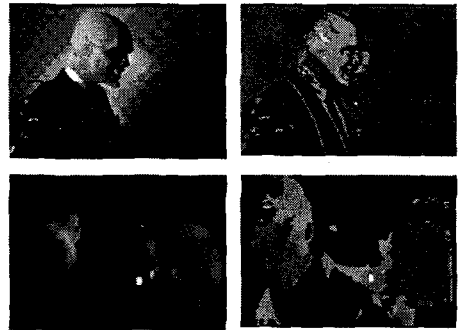


그림 3: 양자화 된 대표 프레임

Hue는 2가지의 집합으로 분리하여 공포와 슬픔을 연상시키는 black과 gray의 한 집합과 즐거움을 연상시키는 yellow, green, brown, pink 및 orange의 집합으로 만들고, 각 집합에 속하는 칼라가 대표 프레임의 전체 픽셀에 대하여 분포하고 있는 비율을 각 집합의 Hue 값으로 정의하였으며, 채도와 밝기는 대표 프레임의 평균 채도와 평균 밝기로서 그 값을 정의하였다.

실험은 감정별로 분류된 데이터의 각 클러스터를 주축 상에 projection하여 분류하는 LDA 방식과 각 클러스터의 평균과 입력 자료와의 거리를 Mahalanobis Distance를 이용하여 측정한 후 분류하는 방식을 사용하였으며, 실험을 위하여 공포, 슬

품, 즐거움을 나타내는 대표 프레임들과 그 나머지 대표 프레임들의 집합들을 생성하였다. 총 3899개의 대표 프레임을 사용하였고, 표 2는 이 집합들에 대한 정보를 나타내고 있다.

표 2: 각 감정별 대표 프레임 집합의 셋 수

	감정이 있는 셋 수
공포	1241
슬픔	130
즐거움	348
감정 없음	2180

실험을 위하여 표 2의 집합을 둘로 나누어 하나의 집합은 감정별로 클러스터링 하는 training set으로 사용하고, 다른 집합은 test set으로 사용하였다.

LDA를 실시하여 공포의 경우 그림 4와 같은 결과를 얻었으며, 가로축은 LDA로부터 얻은 주축이고 세로축은 주축에 projection 된 셋의 수이다. 그림 4는 주축상의 64 지점을 경계로 오른편을 공포 감정이 있는 셋으로 분류하고 왼편을 공포 감정이 없는 셋으로 분류한 경우로서 77.84%의 성공률을 얻을 수 있었다. 그러나 슬픔과 즐거움의 경우는 감정이 있는 셋의 수가 상대적으로 적기 때문에 LDA 주축상에서의 분류가 용이하지 않았다.

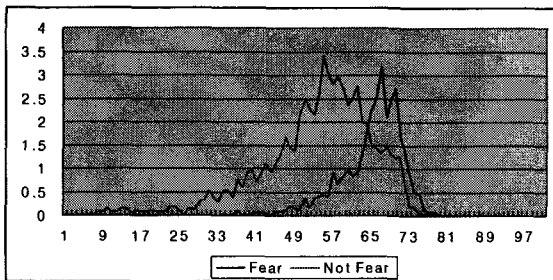


그림 4: LDA 주축상의 셋 분포

Mahalanobis Distance를 이용한 감정 분류 실험은 하나의 집합이 먼저 감정별로 클러스터링 되어 있는 상태에서 다른 집합의 데이터가 하나씩 들어왔을 때 가장 가까운 감정 클러스터로 분류 되는 실험으로서, 공포의 경우는 76.31%의 성공률을 얻을 수 있었으나 슬픔과 즐거움의 경우는 낮은 성공률을 보였다.

두 실험을 통하여 특히 즐거움의 경우는 칼라 정보

와 더불어 다른 정보들의 추가에 의해 구분될 가능성이 높을 것으로 보인다.

## 5. 결론

본 논문에서는 칼라 정보를 이용하여 기본 감정에 관련된 비디오 데이터의 셋을 분류하는 방식을 제안하였다. Black과 gray로 구성된 어두운 칼라, yellow, green, brown, pink, orange로 구성된 밝은 칼라, Saturation 및 Value의 4가지 특징들로 구성된 특징 벡터들을 비디오 셋의 대표프레임으로부터 추출하여, Fisher의 LDA 및 Mahalanobis Distance를 사용하여 분류기를 구성하였다. 공포 감정의 경우 약 77%의 검출률을 얻었으며, 즐거움의 경우는 칼라 정보만으로는 분류가 매우 어렵다는 것을 알았다. 보다 나은 감정에 관련된 비디오 셋을 분류하기 위해서는 칼라 정보이외에도 모션정보 및 음성정보를 결합하는 것이 바람직하다.

## 참고문헌

- [1] R. Picard, *Affective Computing*, MIT Press, 1997.
- [2] R. Picard, "Affective Computing for HCI," *Proc. of HCI'99*, Aug., 1999.
- [3] 강행봉, "감정기반 컴퓨팅" 전자공학회지, 제 28권 12호, pp. 1327-1334, 2001.
- [4] A. Hanjalic and L. Xu, "User-oriented Affective Video Content Analysis," *Proc. IEEE Workshop on CBAIBL'01*, Kauai, HI, pp.50-57, Dec. 2001.
- [5] S. Moncrieff, C. Dorai, and S. Venkatesh, "Affect Computing in Film through Sound Energy Dynamics," *Proc. ACM MM'01*, pp. 525-527, 2001.
- [6] J. Corridoni, A. Del Bimbo, and P. Pala, "Image Retrieval by color semantics," *ACM Multimedia System Journal*, 1998.
- [7] P. Lang, "The emotion probe: Studies of motivation and attention", *American Psychologist*, 50(5), 372-385, 1995.
- [8] E. Goldstein, *Sensation and Perception*, Brooks/Cole, 1999.
- [9] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, Wiley-Interscience, 2002.