비디오 컨텐츠의 시계열 정서 프로파일링

최원희 [1) 2)], 유준희 [2)], 천효선 [2)], 장병탁 [2)]
[1)] 삼성전자 종합기술원
Wonheechoe@gmail.com,
[2)]서울대학교
btzhang@snu.ac.kr

Sequentially Emotional Profiling of video contents

Wonhee Choe, Jun Hee Yoo, Hyo-sun Chun, Byoung-Tak Zhang
Seoul National Univ.

요 약

일반적인 비디오 컨텐츠 분석 툴들은 주로 비디오에 나오는 사물, 장소, 인물 등의 단편적인 개별 컨텐츠를 분석한 정보를 제공하고 있다. 그러나, 실제 사람은 특정 비디오를 시청하고자 할 때 단편적인 컨텐츠를 보고 선택하는 것이 아니라 비디오 컨텐츠에서 전달되는 정서 혹은 분위기에 반응한다. 따라서, 본 연구는 비디오 컨텐츠에서 포함하는 정서를 시간적으로 프로파일링하는 연구로서, 이를 정서 표현을 위한 색조합을 이용하여 컨텐츠에서 표현하고자 하는 정서를 예측하고 이를 비디오의 시계열로 분석하여 프로파일링하고자 한다. 이를 위하여 본 논문에서는 먼저 비디오의 프레임별 주요 색성분 세가지를 추출하고, 이를 바탕으로 각 프레임의 정서를 대표적인 정서공간인 Ekman의 6-정서 공간[1](anger, disgust, fear, happiness, sadness, and surprise)으로 표현하였다. 본 논문의 실험을 위해 비디오들을 정서공간으로 프로파일링하여 각 비디오의 정서 표현구성비 및 시계열 표현 정서변화를 확인할 수 있었다.

1. 서 론

최근 인터넷의 발달에 따라 유튜브와 같은 개인 동영상 저작 및 공유가 활발하게 이루어 지고 있다. 하루에도 수 천 건의 동영상이 인터넷상에 새로 올라오거나 사라지기도 한다. 이와 같이 무수히 많은 데이터에서 원하는 내용을 찾는 것은 간단한 문제가 아니다. 특히, 동영상의 경우 키워드로 찾아진 동영상들에서 원하는 내용을 찾기 위해 검색 결과들을 재생시켜 보면서 확인해야 하는 번거로움이 있다. 영화나 드라마의 경우에도 제목만으로 해당 동영상의 분위기를 예측하기는 쉽지 않다.

일반적인 비디오 컨텐츠 분석 툴들은 주로 비디오에 나오는 사물, 장소, 인물 등의 단편적인 개별 컨텐츠를 분석하여 정보를 제공하고 있다. 실제 우리는 특정 비디오를 시청하고자 할 때 '스포츠카가 나오는 영화' 혹은 '사막 장면이 나오는 영화'와 같이 단편적인 컨텐츠를 보고 선택하는 것이 아니라 '슬픈 영화', '행복한 분위기의 영화'와 같이 컨텐츠에서 전달되는 정서 혹은 분위기를 보고 선택하는 경향이 있다. 이와 같이 사용자의 선택기준과 비디오의 정보제공이 차이를 갖고 있는데, 이를 극복하기 위해, 본 연구에서는 비디오 컨텐츠에서 포함하는 정서의 변화를 시간적으로 분석하여 제시하고자 한다. 이를 위한 첫번째 시도로서, 정서를 표현하는 대표적 수단으로 산업디자인에서 사용되고 있는 조합된 색에 의한 정서표현[2][3]을 참고하여, 영상 혹은 장면에서 표현하고자 하는 정서를 예측하고 이를 비디오의 표현 정서를 시간축으로 나열하여 비디오 특성을 정서로 프로파일링하고자 한다.

영상이 표현하는 정서를 찾기 위해, 본 연구에서는 그림 1의 블록도와 같이 영상에 포함된 색들을 Kobayashi color image scale의 130개 색으로 클러스터링하고, 이를 히스토그램분석을 통해 주요 3가지 색성분을 선정한 다음, 주요 삼색이 표현하는 정서를 찾도록 구성한다. 이와 같이 구성된 알고리즘은 비디오의 프레임별 대표 정서값으로 저장되며, 비디오별 어떤 정서장면들을 포함하는지 한번에 알 수 있도록 하고자 한다.

그림 1. 알고리즘 블록도

2. Emotional Profiling based on 3-Color combinations

정서를 정의하기 위한 공간은 그 목적과 대상에 따라 여러 형태로 제안되고 있다. 그 중에서도 학술적으로 많이 인용되고 있는 것은 Kobayashi의 180 words [2]와 Ekman의 6-정서 공간[1]이 있다. 그림 2는 Ekman이 6-정서 공간을 비언어적으로 실험하기 위해 얼굴표정으로 표현한 예이다.

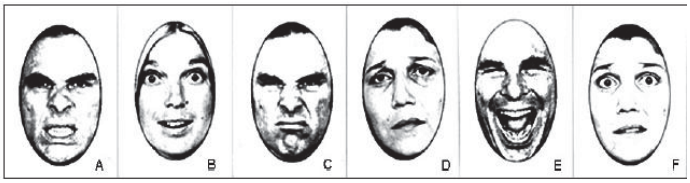

그림 2. 정서의 얼굴 표현: (A) anger, (B) surprise, (C) disgust, (D) sadness, (E) happiness, (F) fear [1]

2.1 Emotional Color Clustering

영상의 색정보를 이용한 분석을 위해서는 8-bit 영상데이터일 경우 2^{8x3}개의 색정보를 분석해야 하며, 노이즈에 의한 정보의 변화에 대해서도 민감하게 반응하기 쉽다. 또한, 2^{8x3}개의 색정보가 모두 정서에 유의미한 정보를 갖고 있는지도 명확하지 않다. 이에 대해 Kobayashi는 정서적으로 유의미한 color scale을 재 정의하였으며 이를 130 hue and tone matrix로 표현하였다[2]. 본 연구에서는 정서적으로 유의미한 색을 추출하기 위해 16777216가지 RGB 색(RGB_j)을 hue와 tone 130가지 색($HTIndex_{RGBj}$)으로 식 (1)과 같이 유클리디안 거리를 이용하여 가장 가까운 색으로 분류한다. 연산의 단순화를 위해 look-up-table(LUT) 형태로 생성하여 입력 영상에 대한 화소별 RGB 값에 대해 130개의 색정보로 매핑되도록 한다.

$$HT\ Index_{RGB_j} = \arg\min_{1\le i\le 130}\left(dist_{ij}\left(RGB_i, RGB_j\right)\right) \qquad (1)$$

2.2 3-Color Selection with Histogram

각 프레임별 주요 색을 찾기 위해 130개 색에 대해 알고리즘(1)과 같이 히스토그램 분석을 수행하여 가장 많은 색정보 3가지를 선정한다.

$S = \{s_i: \text{the number of } HueAndTone_i \text{color pixel in frame}\}$
FOR $rank = 1$ to 3 DO
  $\text{Dominant color}_{\text{rank}} = \arg\max_i(S)$
  $S := S - \text{Dominant color}_{\text{rank}}$
ENDFOR

알고리즘 1 최빈값 3개를 찾는 방법

2.3 6-Emotion Mapping

임의의 색조합이 정서를 표현한다는 가정은 심리학, 산업공학, 디자인, 칼라 사이언스 분야에는 공공연한 사실처럼 여겨지고 있다. Kobayashi는 180개의 정서 형용사를 1170개의 세가지 색조합으로 표현하고 이를 15개의 대표 그룹으로 묶었고[2], 몇몇 연구에서는 이를 실제 정지영상에서 정서를 정의하는 방법으로 활용하기도 한다[4][5]. 그러나 산업공학이나 디자인과 같이 의도하여 장면을 재구성한 것이 아닌 일반 영상에서 미묘한 180개의 형용사를 정의하는 것은 어려우므로, 본 연구에서는 6개의 대표감성과 3-색 조합과의 연관성을 연구한 Pos와 Armytage의 연구결과를 활용하고자 한다[3]. 3가지 주요색 조합이 해당하는 정서를 아래와 같은 정서별 색조합에 해당하는지 찾고 없는 경우 2가지 색조합이 해당하는 정서가 많은 경우를 선택하는 방식으로 동작한다. 동작방식은 식 (2)과 같이 표현된다.

$$\text{OutE}_f = \arg\max_{1\le i\le 6}\left(\sum_{m=1}^{3}\frac{n_{E_i,m}}{10^{3-m}}\right) \qquad (2)$$

E_i: Emotions ($1 \le i \le 6$: Surprise, Sad, Fear, Happy, Disgust, Angry)

m: 하나의 색 조합에 대해 3가지 주요색 조합 과 일치하는 색의 수

$n_{E_i,m}$: E_i의 색 조합들 중 3가지 주요색 조합과 m개 색이 일치하는 조합의 수

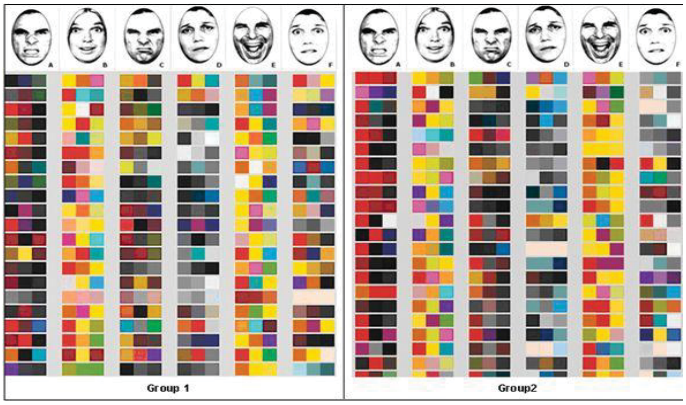

그림 3. Pos와 Armytage의 6개 정서에 따른 3색 조합 결과[3]

3. 실험 및 결과

3.1 실험 설계

동영상에 대한 정서 프로파일링 실험을 수행하기 위해 10분 내외의 동영상에 대해 실험하고, 영상은 5 프레임 간격으로 샘플링하여 정서를 분석하도록 한다.

3.2 실험 결과

실험에 사용된 동영상은 총 3편으로 그림4와 그림 5와

같은 정서 분포를 갖는 것으로 나타났다. 그림 4와5는 각각의 동영상 전체의 정서변화를 프로파일링한 다음 동영상 내에 대표정서를 포함하는 장면의 비율을 파이 차트로 표현한 것이며, 그림 5의 (b)는 각 정서별 동영상의 프로파일링한 것을 Happy와 Angry에 대해서만 표현한 것이다.

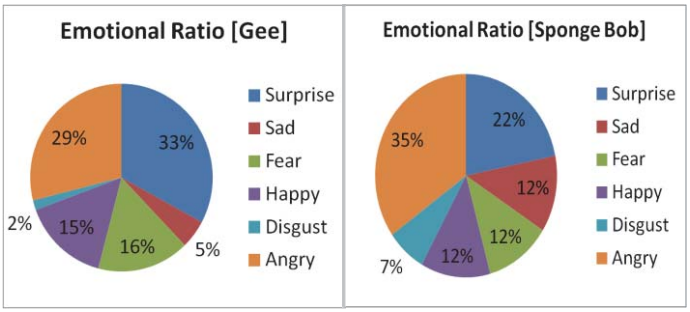

그림 4. 뮤직비디오"Gee"와 애니메이션"스폰지밥"의 프레임별 정서 표현 분석 결과

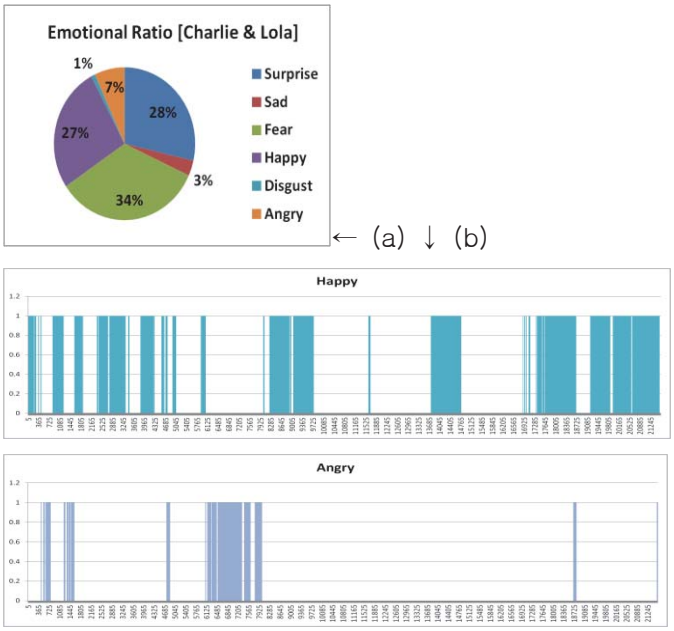

그림 5. 애니메이션 "Charlie & Lola"에 대한 정서변화 프로파일링

4. Discussion and Future Works

본 연구는 동영상에서의 단순 컨텐츠 분석에서 벗어나 비디오 컨텐츠의 정서를 알아내고자 하였다. 그 첫번째시도로서 본 논문에서는 영상이 포함하는 색정보들로부터 표현하고자 하는 정서를 예측하고자 하였다. 그러나 Ekman의 6-정서공간은 긍정적 단어 대비 부정적 단어 비중이 높아 고른 분포의 색을 포함하는 영상에서는 부정적인 정서로 할당될 확률이 높게 나오는 단점이 있다. 또한, 해당 정서공간의 정서표현이 동영상의 컨텐츠에 대한 정서로 표현하기 적절한지에 대해서도 추가적인 연구가 필요하다. 즉, 우리가 일반적으로 해당 비디오 컨텐츠를 보고 느끼는 정서인 Exciting의 경우 Surprise와 Angry의 조합을 Exciting으로 해석될 수 있다.

향후 연구에서는 실제 컨텐츠 내용에 따른 사용자 반응에 적절한 정서공간 선정 및 사용자 실험을 통해 비디오 컨텐츠 분석만으로 정확도 높은 사용자 정서 예측 모델을 연구하고자 한다.

Acknowledgement

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(No. 2012-0005643, Videome), 정부(지식경제부)의 재원으로 한국산업기술평가관리원의 지원(10035348, mLife) 및 교육과학기술부의 BK21-IT 프로그램에서 일부 지원되었음.

참고문헌

[1] P. Ekman and W. V. Friesen, *Manual for facial action coding system*, Palo Alto: Consulting Psychologists Press (1978)

[2] Osvlado da Pos and Paul Green Armytage, "Facial Expressions, Colours and Basic Emotions," Colour: Design & Creativity 1(1) 2, 1-20. (2007).

[3] S. Kobayashi, "Aim and method of the color image scale," Color Res. Appl., 6(2): 93-107, 1981.

[4] M. Solli and R. Lensz, "Color semantics for image indexing," in Proceeding CGIV 2010, 5th European Conference on Colour in Graphics, Imaging, and Vision, pp.353-358, 2010

[5] 신윤희, 김은이, "감성기반 영상검색을 위한 확률적 감성모델 구현," 정보과학회논문지: 소프트웨어 및 응용 제 38권 제11호(2011, 11) pp. 579-590