저급 특징들로부터 비디오 셧의 특성 분석

박현재°, 강행봉
가톨릭대학교 컴퓨터 공학과
{hyunjapark°, hbkang}@catholic.ac.kr

Analysis of video shots' characteristics using low-level features

Hyun-Jae Park° and Hang-Bong Kang
School of Computer Science, The Catholic University of Korea

요 약

본 논문에서는 비디오 데이터에서 비디오 셧이 가지는 저급 특징들로부터 감정에 관련된 특징을 검출하기 위하여 비디오 셧의 특성을 확률적 분포를 이용하여 모델링 하는 방식을 제안한다. 제안한 방법을 통해 감상자가 감정을 느끼게 하는 부분의 비디오 셧을 검출할 수 있는 방법에 대하여 기술한다. 특징값과 감정과의 관계, 시간의 흐름과 감정과의 관계를 통계적으로 분석하여 모델링 함으로써 감정 검출이 가능하다는 것을 확인하였다.

1. 서 론

비디오 데이터를 분석하는 것은 다양한 분야에서 응용될 수 있는 연구 주제이다. 비디오 데이터는 작은 주제를 보여주는 장면(Scene)과 이런 장면을 구성하는 셧으로 이루어진다. 셧은 연속된 프레임의 집합으로 비디오 데이터를 구성하는 요소 중 의미를 전달하는 가장 작은 단위이다. 이러한 셧의 특징은 비디오 데이터의 분석을 위한 기초적인 단위가 된다.

Vasconcelos and Lippmann[1]은 비디오 셧이 갖고 있는 정보인 셧의 길이나 움직임의 확률 분포를 모델링 하여 베이시안 프레임워크를 사용하여 셧 경계 검출을 시도하였다. 또 셧에 대한 정보를 이용하여 비디오 장르로 구분하였다. 또, Moncrieff et al.[2]는 비디오 셧의 사운드 정보로부터 감정 분류 방법을 제안하여 공포 감정을 갖는 장면을 비디오 데이터로부터 추출하였다. Hanjalic and Xu[3]는 비디오 셧의 모션 및 사운드 정보를 이용하여 사용자의 감정 곡선을 생성 및 분석하였다.

비디오 셧의 저급 특징들로부터 사용자의 감정 상태를 추출하는 것은 매우 어려운 작업이다. 하지만, 비디오의 특징으로부터 감정에 관련된 특징을 찾아내고 확률적인 분포를 사용하여 통계적인 모델을 생성, 감정 검출에 적용하면 저급 특징과 고차원의 감정과의 매핑을 예측할 수 있다. 이것은 강건한 결과를 얻기 위한 바람직한 방법 중의 하나이다.

본 논문에서는 비디오 셧이 가지고 있는 저급 특징(low level features)들인 컬러, 모션 및 셧의 길이의 확률 분포를 모델링 하여 감정 비디오 셧을 검출하는 방법을 제안한다. 제 2절에서는 비디오 셧의 특징에 대하여 기술한다. 제 3절에서는 비디오 셧의 특징을 감정 별로 모델링 하는 기법에 대하여 설명한다. 제 4절에서는 감정 추출 방법을 제공하는 베이시안 분류법에 대하여 설명한다. 제 5절에서는 실험 결과를 분석한다.

2. 특징 모델링

본 논문에서 시도하는 감정 검출 기법은 비디오 데이터에서 추출할 수 있는 특징들을 확률적인 모델로 구성하고, 통계적인 분석을 통해 감정을 발생하는 셧과 그렇지 않은 셧을 판단하는 작업이다. 저급 특징들을 사용하여 확률 모델을 구성할 때 이용하는 통계적인 모델을 두 가지로 나눠볼 수 있다. 하나는 시간에 따른 감정발생 여부를 모델링 하는 것이고, 다른 하나는 셧 내에서의 특징값을 모델링 하는 것이다.

2.1. 시간에 따른 감정 변화 모델링

비디오 데이터에서 발생하는 감정은 대체로 시간에 따라 변화한다. 일정 시간 간격을 두고 발생하며, 한 번 발생한 감정은 일정 시간 지속한다. 이러한 성질을 모델링하기 위하여 몇 가지 확률 모델을 도입하고, 통계적인 분석을 시도한다. 감정이 발생할 확률은 시간에 따라 변화하므로 Erlang, Weibull, Exponential 과정 등을 사용할 수 있다. 이러한 시간에 따른 확률 모델들을 통계적으로 분석하여 모델링하고, 이를 바탕으로 감정이 발생하는 부분들을 찾아낸다.

이러한 확률 모델들을 이용해 학습 데이터를 가장 잘 나타내는 모델 파라미터를 찾아냄으로써 시간에 따라 변화하는 감정 영역을 예측할 수 있다. 모델 파라미터를 추정하기 위하여 MLE(Maximum Likelihood Estimation)가 사용된다.

2.2. 셧의 특징값 모델링

비디오 데이터를 이용하여 감정 분류를 구현하기 위해서는 가장 먼저 비디오의 특징값을 계산하는 과정이 필요하다. 비디오의 저급 특징으로는 컬러 정보와 모션 정보, 셧 내의 움직임 크기 그리고 셧의 길이 정보를 사용한다.

컬러 정보를 감정 레벨에서 표현하기 위하여 셧 내의 채도(Saturation), 밝기(Luminance Intensity), Single Dominant

Colour를 검출한다. Single Dominant Colour는 셋 내에서 절대적으로 많은 영역을 차지하는 색이 있는지 나타내는 특징이다.

카메라 모션 검출을 위해서는 화면의 영역을 세부 영역으로 나누고 PAN, TILT, ZOOM의 템플릿과 비교하는 템플릿 매칭 방법을 사용한다[4].

셋 내의 움직임 크기(Shot Activity)는 프레임 차이로 표현되는데, 이것은 카메라 모션이 없다고 판단될 때 계산된다. 카메라의 움직임이 없을 때, 셋 내의 프레임 차이를 구하면 객체의 움직임의 크기를 구할 수 있다. 그리고 셋 길이는 영화의 성격을 나타내주는 간결하면서도 좋은 결과를 보이는 널리 이용되는 특징이다.

본 논문에서 사용되는 특징 벡터 세트는 다차원의 특징 공간(Feature Space)에 존재하며, 정밀한 분포를 표현하기 위하여, GMM(Gaussian Mixture Model)을 이용한다. GMM의 파라미터를 구하기 위하여 EM(Expectation Maximisation)을 사용하여 모델 파라미터를 추정한다.

3. 통계적 모델을 이용한 베이시안 분류법

앞 장에서 설명한 특징 모델링을 이용하여 학습 샘플들의 분포를 모델화하고 계산된 모델을 기반으로 테스트 샘플의 확률을 나타내어, 베이시안에 이용한다.

i번째 셋에서 감정이 발생하고, $i-e$에서 비감정에서 감정으로의 천이가 발생한 경우 $i-r$에서 감정에서 비감정으로의 천이가 발생할 확률은 다음과 같다.

$$P(S_i=1 \mid S_{i-e-1}=0, S_{i-e}=1, S_{i-r-1}=1, S_{i-r}=0, D_i) \quad (1)$$

이 때 D_i는 i번째 셋의 특징 벡터이다. $s_i=1$은 i번째 셋에서 감정이 발생한 경우이며, $s_i=0$의 경우 반대로 감정이 발생하지 않은 경우를 나타낸다.

위의 과정과 동일한 조건하에 i번째 셋이 비감정일 확률을 구해보면 다음과 같다.

$$P(S_i=0 \mid S_{i-e-1}=0, S_{i-e}=1, S_{i-r-1}=1, S_{i-r}=0, D_i) \quad (2)$$

두 사후 확률 (1), (2)을 이용하여 베이시안 분류법을 적용하면 식 (3)을 구성할 수 있다.

$$\frac{P(S_i=1 \mid S_{i-e-1}=0, S_{i-e}=1, S_{i-r-1}=1, S_{i-r}=0, D_i)}{P(S_i=0 \mid S_{i-e-1}=0, S_{i-e}=1, S_{i-r-1}=1, S_{i-r}=0, D_i)} \quad (3)$$

S_{i-e}, S_{i-e}, S_i, S_{i-r-1}들이 모두 독립적이라 가정하고, 식 (3)을 정리하면 다음과 같다.

$$\log\frac{P(S_{i-e-1}=0, S_{i-e}=1 \mid S_i=1)}{P(S_{i-e-1}=0, S_{i-e}=1 \mid S_i=0)} + \log\frac{P(S_{i-r-1}=1, S_{i-r}=0 \mid S_i=1)}{P(S_{i-r-1}=1, S_{i-r}=0 \mid S_i=0)} + \log\frac{P(D_i \mid S_i=1)}{P(D_i \mid S_i=0)} + \log\frac{P(S_i=1)}{P(S_i=0)}$$

식(3)의 각 항을 아래와 같이 간단한 표현으로 나타낼 수 있다.

$$g(e)+h(r)+f(D_i)+C \quad (4)$$

$$g(e)=\log\frac{P(S_{i-e-1}=0, S_{i-e}=1 \mid S_i=1)}{P(S_{i-e-1}=0, S_{i-e}=1 \mid S_i=0)}$$

$$h(r)=\log\frac{P(S_{i-r-1}=1, S_{i-r}=0 \mid S_i=1)}{P(S_{i-r-1}=1, S_{i-r}=0 \mid S_i=0)}$$

$$f(D_i)=\log\frac{P(D_i \mid S_i=1)}{P(D_i \mid S_i=0)}$$

$$C=\log\frac{P(S_i=1)}{P(S_i=0)}$$

이를 이용하여 분류식을 다음과 같이 정리한다.

$$\begin{cases} S_i=1 & if \quad f(D_i) \geq -(g(e)+h(r)+C)=T(r,e) \\ S_0=0 & otherwise \end{cases} \quad (5)$$

$g(e)$는 비감정 영역에서 감정 영역으로의 천이가 일어나고 e개의 셋이 지나간 후 감정 영역이 나타날 확률에 대한 시간 분포 함수이다. $h(r)$은 감정 영역에서 비감정 영역으로의 천이가 일어나고 r개의 셋 이후에 감정 영역이 나타날 확률에 대한 시간 분포 함수이다.

이러한 시간 분포 모델을 도입함으로써 얻을 수 있는 이득은 일정 시간 동안 연속적으로 나타나는 감정 영역을 끊지 않고 연속적으로 검출해 낼 수 있다는 점이다.

4. 실험 결과 및 분석

실험은 영화 데이터 14개를 이용하여 감정 분류를 시도하였다. 각 영화는 10명중 7명이상이 같은 감정을 가질 때 감정이 있는 셋으로 구분하였다.

그림 1과 2는 감정 발생 셋의 모델링 시에 사용하는 가우시안 컴포넌트의 개수에 따라 변화하는 Correct Acceptance Rate와 FAR(False Acceptance Rate)의 그래프이다. 가우시안 컴포넌트의 개수가 높아짐에 따라 FAR이 낮아짐을 볼 수 있다. 컴포넌트의 개수가 많아짐에 따라 Correct Acceptance 역시 낮아지지만 FAR의 변화율보다 훨씬 작은 것을 확인할 수 있다. 그리고 실선과 점선으로 표시된 각 그래프는 비감정 셋을 모델링하기 위한 가우시안 컴포넌트의 개수이다. 그림 1과 2를 보면 비감정 셋을 모델링하기 위해 적합한 가우시안 컴포넌트의 개수는 4~5임을 알 수 있다.

그림 3은 식 (4)에서 각 항을 사용하였을 때의 결과이다. 모든 항을 사용하였을 때 그렇지 않은 경우보다 False Positive rate이 낮은 것을 볼 수 있다. 다른 경우와 달리 모든 항을 사용한 경우 False positive rate이 50% 미만에 분포하는 것을 볼 수 있다. 이것은 $g(e)$와 $h(r)$의 조합으로 인해 $T(r,e)$의 값이 상승, 하강을 반복하게 되어 나타난 결과이다. 다른 경우

의 분포를 보면 Correct Positive-False Positive 평면의 좌하단과 우상단에 주로 분포함을 확인할 수 있다. 이러한 결과는 분류 결과가 한 쪽으로 치우쳐 모든 비디오 셋이 감정 발생 셋으로 분류됐든지 아니면 모두 비감정 셋으로 분류된 경우 나타나는 결과이다. 결국 $g(e)$와 $h(r)$의 조합으로 이와 같은 현상을 피할 수 있음을 알 수 있다.

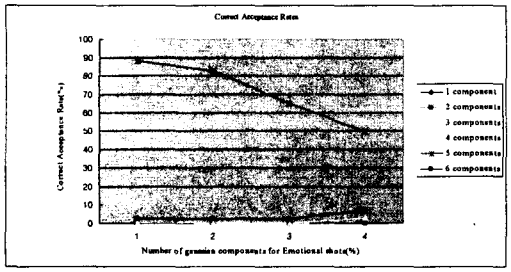

그림 1 : 가우시안의 개수에 따른 비감정 영역 모델 결과

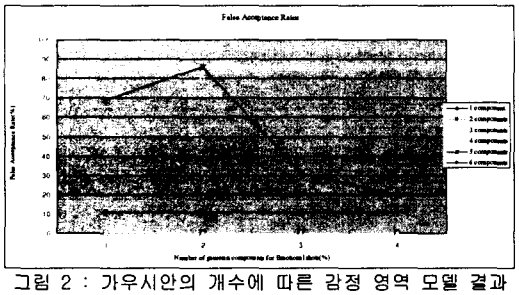

그림 2 : 가우시안의 개수에 따른 감정 영역 모델 결과

, 그림 4는 분류에 사용되는 특징들을 변화시켜 얻은 결과이다. 셋 길이와 셋 움직임만을 이용하여 11개의 특징을 구성한 경우, 그 것에 컬러 정보를 추가하여 14개의 특징을 구성한 경우 그리고 카메라 모션을 더하여 17개의 특징을 구성한 경우의 결과를 볼 수 있다. 17개의 특징을 사용하였을 때 다른 경우보다 더 좋은 결과를 얻을 수 있었다.

5. 결론

본 논문에서는 비디오 데이터에서 감정을 발생시키는 셋을 찾아내기 위한 방법으로 통계적인 모델을 이용하였다. 시간의 변화에 따라 변화하는 감정과 셋 내에서의 특징값에 대하여 통계적인 분석을 한다. 그것을 바탕으로 모델을 생성하고 테스트 비디오 데이터의 감정을 검출하였다. 이러한 방법을 시도하여 감정 유발 비디오 셋을 분류한 결과 감정의 분류 가능성을 확인하였다.

비디오 데이터에서의 감정 분류를 위해서 앞으로 음향 효과의 특징을 도입하면 좀 더 좋은 결과를 얻을 수 있을 것으로 보인다.

참고 문헌

[1] Nuno Vasconcelos and Andrew Lippman, "Statistical Models of Video for Content Analysis and Characterization", IEEE Trans. Image Processing, vol. 9, pp. 3-19, Jan. 2000

[2] S. Moncrief, C. Dorai and S. Venkatesh, "Affect Computing in Film through Sound Energy Dynamics", Proc. ACM MM'01, pp525-527, 2001

[3] A. Hanjalic and L. Xu, "User-oriented Affective Video Content Analysis", Proc. IEEE Workshop on CBAIBL '01, Kauai, HI, pp50-57, Dec 2001

[4] Sangkeun Lee; Hayes, M.H., III, "Real-time camera motion classification for content-based indexing and retrieval using templates", Proceedings. (ICASSP '02). IEEE International Conference , Volume: 4 , 2002 , pp.3664 -36

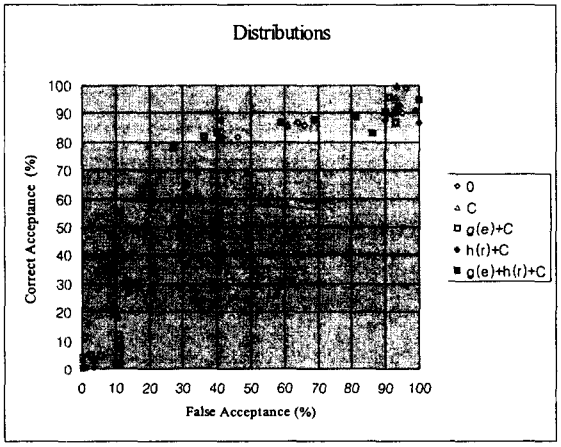

그림 3 : 각 항별로의 결과 그래프

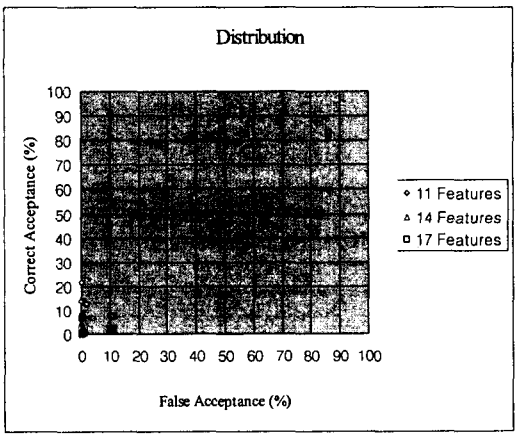

그림 4 : 특징의 변화에 따른 결과 분포