

인터랙티브 비디오 검색을 위한 EOG 기반 장면 전환 검출*

이충연¹, 이범진², 장병탁^{1,2}

{cylee, bjlee, btzhang}@bi.snu.ac.kr

¹서울대학교 뇌과학협동과정, ²서울대학교 컴퓨터공학부

Electrooculogram-based Scene Transition Detection for Interactive Video Retrieval

Chung-Yeon Lee¹, Beom-Jin Lee², Byoung-Tak Zhang^{1,2}

¹Interdisciplinary Program in Neuroscience, Seoul National University, ²School of Computer Science and Engineering, Seoul National University

요 약

기존의 비디오 검색 방법들은 관련 주석이나 영상 정보에 기반하며 사용자의 반응과 관련하여서는 많은 정보를 활용하고 있지 않다. 비디오 시청시 사용자의 뇌신호나 시선추적 정보 등의 인지적 반응을 이용하여 연속적인 비디오 스트림의 각 부분에 대하여 사용자들이 나타내는 관심이나 감성 정보를 추출한다면 보다 인터랙티브한 비디오 데이터 검색 및 추천이 가능하다. 본 논문에서는 비디오를 시청하는 사용자의 안구전도(electrooculogram)를 기록한 후, 장면 전환이 발생한 부분에서의 사건관련전위 분석을 통해 해당 부분에서 나타나는 특징적 반응을 찾고 이에 대한 인지적 해석을 도출했다. 실험 결과 장면 전환 이후 200~700ms 부분에서 P300 성분과 유사한 피크가 발생하는 것을 확인하였으며, 이러한 결과는 장면 전환에 따른 피험자의 비디오 내용 인지에 대한 의도 불일치 및 주의력 증가로 해석된다.

1. 서 론

최근 압축 기술과 통신 기술의 급격한 발달에 따라 멀티미디어 데이터의 활용이 폭발적으로 증가하고 있으며, 멀티미디어 데이터를 효율적으로 전송, 저장, 관리 및 검색하는 기술이 핵심 기술로 대두되고 있다. 그 중에서 디지털 비디오 정보 검색의 경우 사용자가 원하는 정보를 표현할 수 있는 사용자 인터페이스(user interface) 기술과 원하는 정보를 사용자에게 신속하고 정확하게 보여주는 기술(rapid browsing)의 필요성이 증대하고 있다 [1].

현재 디지털 비디오 데이터의 검색 방법은 크게 두 가지로 나눌 수 있다. 하나는 주석 기반 검색이고 다른 하나는 내용 기반 검색이다. 전자의 방법은 개발이 간편하고, 개인 및 소셜 네트워크(social network)로부터 추출되는 사용자의 기호도나 이용 패턴 등 다양한 정보들을 검색에 활용할 수 있다는 장점으로 인하여 많이 사용된다. 하지만 이런 주석 기반의 영상 검색은 수작업으로 키워드를 입력하는데 따르는 불편함, 검색의 비효율성, 검색 대상에 대한 적합한 키워드 선택 등의 제약으로 인하여 문제가 되고 있다. 반면 내용 기반 검색은 디지털 비디오 데이터에서 내용으로 기술되는 특징 데이터를 자동

로 추출하여 이를 기반으로 검색을 수행하는 방법이다 [2].

디지털 비디오의 내용 기반 검색을 위하여 가장 일반적으로 사용되는 정보는 비디오를 구성하는 시퀀스가 가지는 영상 정보들로 색상(color), 모양(shape), 질감(texture) 등의 속성이 주로 이용된다. 이러한 영상 정보는 주로 비디오를 장면 분할할 때에 사용되며 이를 통하여 구조적인 비디오 브라우징을 할 수 있다. 비디오를 샷(shot)으로 구분하는 작업을 비디오 분할(video segmentation)이라고 하며, 비디오 분할을 위해 장면의 전환점인 컷을 검출하는 작업을 컷 검출(cut detection)이라고 한다. 비디오 분할 기술은 장면 전환 효과에 의해 발생하는 샷 사이의 경계를 검출하는 방법으로 샷은 비디오 분할의 기본 단위로 사용된다.

하지만 이러한 검색 방법들은 해당 비디오 데이터 내에서 연속적으로 변하는 부분적인 이야기(story) 정보나 이에 대한 사용자의 감성 변화 정보를 추출하지 못하고, 비디오 데이터의 대표적인 주석(장르, 배우, 별점 등)이나 영상 정보 그 자체만을 사용하는데 그치고 있다. 색상이나 조명 변화로부터 감성 평가 모델을 구축하여 비디오 데이터의 감성 정보를 추출한 연구들[1, 3]이 있었지만, 이 또한 사용자 정보가 아닌 비디오 영상 자체로부터 추출한 정보이기 때문에 한계가 있는 것으로 판단된다.

만약 뇌전도(electroencephalogram, EEG) [4-5], 시선 추적(eye-tracking) 정보 [6], 심전도 [7] 등 사용자의 정보 처리 과정 및 집중도, 감성 등을 반영하는 인지적 반응을 이용하여 사용자 정보를 추출한다면, 연속적인 비디

* 이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(No. 2012-0005643, Videome, No. 2012.0005801, BrainNet), 정부(지식경제부)의 재원으로 한국산업기술평가관리원의 지원(No. 10035348, mLIFE) 및 교육과학기술부의 BK21-IT 프로그램에서 일부 지원되었음.

오 데이터의 각 부분에 대하여 사용자들이 나타내는 관심이나 감성 정보를 보편적, 주관적으로 나누어서 추출해냄으로써 보다 사용자 인터랙티브(user-interactive)한 비디오 데이터 검색 및 추천이 가능할 것이다.

본 논문에서는 비디오를 시청하는 사용자의 안전도(electrooculargram, 이하 EOG [8]) 데이터로부터 장면 단위의 비디오 장면 전환에 따른 특징을 분석하였다. 장면 전환 효과에 대한 사용자 반응은 사용자가 비디오를 주의 깊게 시청하고 있는지와 해당 부분에서의 이야기 정보에 대한 분류 여부를 파악하는데 중요하며, 이를 이용한 비디오 분할은 비디오 신속 브라우징 및 주요 스토리 파악, 그리고 오브젝트의 등장이나 모션 등 다른 이벤트에 의한 사용자 반응 분석에도 활용될 수 있다.

2. 연구 내용 및 방법

본 연구는 비디오를 시청할 때의 피험자 EOG를 기록하고, 비디오의 장면 전환 시점에 대하여 기록된 EOG를 장면 전환이 없는 부분에서 기록된 것과 비교하여 나타나는 특징을 찾는 순서로 수행된다. 이때 각 장면 전환 시점은 장면 전환이 없는 다른 비디오 부분과 비교하여 특정 자극으로 작용함으로써 특징적인 EOG 신호 변화를 일으킨다는 가설을 가지고 연구를 진행하였다.

본 논문에서는 장면 전환에 따른 EOG에서의 특징을 추출하기 위해 사건관련전위(event-related potentials, 이하 ERP)를 분석하였다. ERP는 시각이나 청각을 통해 특정 자극이 주어졌을 때 그와 관련하여 일정 시간 동안 나타나는 뇌의 전기적 활동을 가리킨다 [9]. ERP는 EEG에서 나타나는 특징 분석에 주로 사용되지만, 본 논문에서는 EOG 변화 또한 특정 뇌신호에 의해 영향을 받는다는 가정을 두고 ERP를 이용한 분석을 시도하였다.

ERP는 특정 자극이 반복적으로 제시되는 동안 나타나는 시간 한정적(time-locked)인 EEG 성분을 자극 제시 시점부터 일정한 구간으로 잘라 평균을 산출함으로써 측정한다. 구간 내에서 시간적으로 사건에 묶이지 않는 무작위적 EEG들은 평균 산출 과정에서 감소하고, 신호 대 잡음 비(signal-to-noise ratio)의 증가에 따라 자극에 의해 발생하는 ERP는 더욱 뚜렷이 남게 된다. 또한 ERP는 시간해상도가 뛰어나 자극의 제시와 동시에 그 처리 과정을 msec(1/1000초) 단위로 직접 그리고 연속적으로 관

찰할 수 있기 때문에 순간적으로 변하는 비디오 장면 전환에 대한 EEG 변화 양상을 분석하는데 적합한 방법이다 [10].

EOG 획득을 위한 실험에는 만 20~28세(평균 24.0 ± 2.7 세)의 대학생 8명이 참여하였으며, 각 피험자는 그림 1과 같이 양쪽 관자놀이 부분에 각각 EOG 전극(bi-channel, Ag/AgCl)을 부착하고, 27분 길이의 자막이 출력되지 않는 외화 시트콤(Friends, Season 10, Episode 5)을 시청한다. 이때 피험자가 따로 수행해야 할 과제는 없으며, 비디오 재생 시작과 동시에 피험자의 EOG가 1000Hz로 기록된다. 기록에 사용된 장비는 Neuroscan사의 SynAmps2 Amplifier이며, 데이터 분석에는 MATLAB과 EEGLAB Toolbox가 사용된다.

비디오 시청 중 획득한 피험자별 EOG는 먼저 기저선 제거(baseline removal)를 수행한 후, 비디오에서 장면 전환이 발생한 10개 시점에 대한 epoch (-500~1000ms)을 추출한다. 또한 비교 데이터로 사용하기 위해 각 장면 전환 시점에서 2초 이전의 epoch을 같은 범위만큼 추출한다. 이후 모든 피험자들의 epoch을 장면전환 여부에 따라 분류하여 두 개의 데이터 집합으로 만들고, 각 데이터 집합으로부터 ERP를 측정한다. 이때 각 피험자 데이터에서의 개인차를 극복하기 위해 rectangular (boxcar) moving average를 이용하여 스무딩(smoothing, cross adjacent trials)하였다.

3. 실험 결과

장면 전환 시점에 대한 ERP (transition)와 장면 전환이 없는 비교 데이터에 대한 ERP (normal)를 비교한 결과를 그림 2에서 제시하였다. 그림에서 x축은 장면 전환이 발생한 시점인 0 ms을 기준으로 한 epoch에서의 시간을 millisecond (ms) 단위로 나타내며, y축은 해당 시간에서의 EOG 전위를 microvolt (μV) 단위로 나타낸다.

그림 2에서와 같이 장면 전환 이후 200ms부터 700ms 사이에서 두 ERP가 다른 양상을 나타내는 것을 알 수 있다. 즉, 장면 전환 시점에서의 ERP는 그렇지 않은 일반적인 다른 부분과 비교해서 보다 큰 전위를 발생하였으며($p < 0.01$), 반면 200ms 이전과 700ms 이후에는 두 전위가 큰 차이를 나타내지 않음을 발견하였다. 그림 3은 각각의 ERP를 시행 단위(trial) 별로 도시한 결과이다. 마

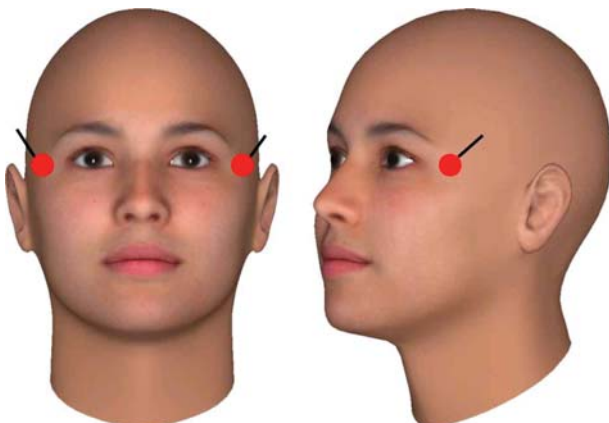


그림 1. EOG 전극 부착 위치

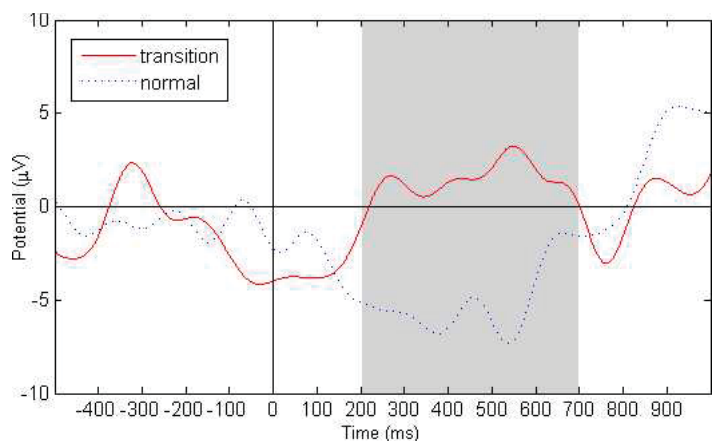


그림 2. 장면 전환 ERP와 일반 EOG 전위 비교 결과

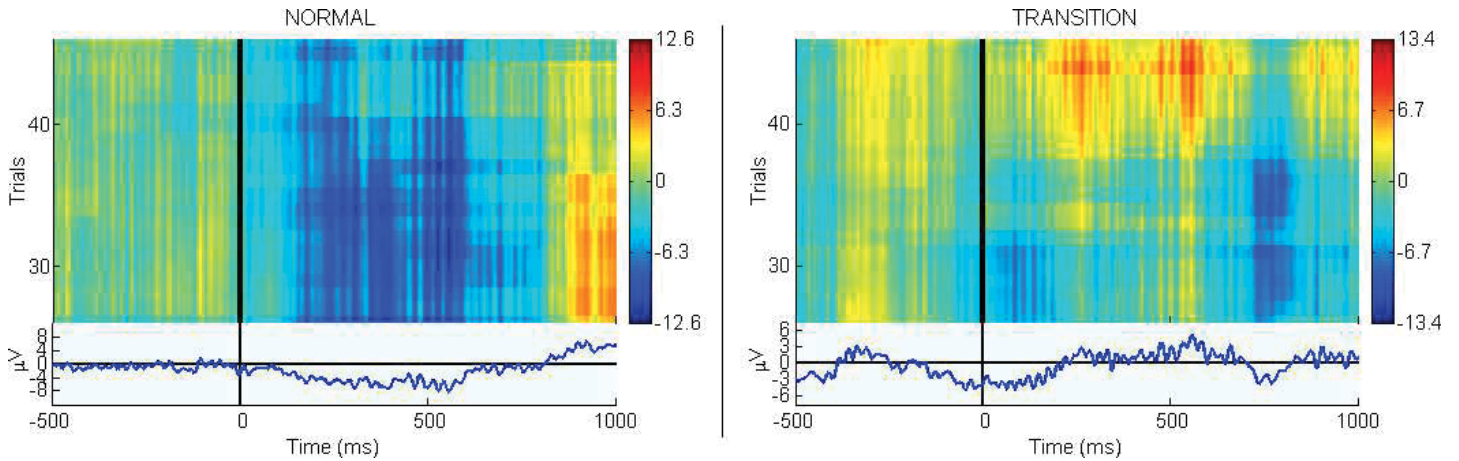


그림 3. 일반 EOG와 장면 전환 부분 EOG의 시행 단위별 ERP 출력 결과

참가지로 장면 전환이 없는 부분의 EOG (좌측)와 비교할 때, 모든 시행에서 장면 전환이 발생한 EOG의 ERP (우측)가 더 높은 전위를 나타낸 것을 확인할 수 있다.

보통 ERP는 다양한 의미를 지니는 여러 가지 피크(N100, N200, P300, N400, P600, P800 등)로 구성되는데, 특히 이러한 피크 중 P300에 해당하는 부분이 뇌의 정보처리 기전과 관련하여 가장 많이 연구되어왔다. P300이란 자극제시 후 약 300ms 지점에 나타나는 양(positive, 상향)의 피크를 의미하며, 보고된 선행연구에 의하면 P300은 정보처리과정 중 자극에 대한 주의력, 자극인지, 기억탐색, 불확실감의 해소 등을 반영한다. 따라서 그림 2와 3의 결과는 시각적 자극이 인간의 의도와 일치하지 않을 경우 증가하는 P300의 특성을 잘 나타내며, 이는 또한 피험자의 주의력 증가에도 관련이 있는 것으로 볼 수 있다. 즉, 비디오 시청 중 장면 전환이 발생하는 순간 피험자가 인지하는 이야기 흐름에 대한 불확실성이 증가하고, 다음 화면에 대해서 의도 불일치가 발생하며, 주의력이 증가하기 때문에 장면 전환이 없는 부분의 EOG와 비교해서 200~700ms 사이에서 P300 피크가 발생하는 것으로 해석 가능하다.

4. 결론 및 향후연구

본 논문에서는 비디오를 시청하는 사용자의 EOG를 기록하고 여기서 장면 전환이 발생한 부분에서의 ERP를 그렇지 않은 부분과 비교 분석하였다. 실험 결과 장면 전환시 ERP에서 P300 피크가 발생하는 것을 확인할 수 있었으며, 이는 장면 전환에 따른 피험자의 비디오 인지에 대한 의도 불일치 및 주의력 증가로 해석된다.

연구 내용은 비디오를 시청할 때의 사용자의 인지적 반응으로부터 비디오 스트림의 특정 부분에 대한 사용자 정보를 추출함으로써 인터랙티브한 비디오 데이터 검색 및 추천에 활용하는 연구의 일환으로, 이 논문에서는 장면 전환에 대한 EOG 전위의 특징 추출 및 해석 단계에서 논의를 마치고, 이를 이용한 장면 전환 탐색이나 다른 인지적 사용자 반응 분석은 향후 연구에서 수행할 계획이다. 특히 시선 추적(eye-tracking) 정보와 EOG를 함께 분석한다면, 보다 구체적이고 의미 있는 실험 결과를 얻을 수 있을 것이라고 예상된다.

참고문헌

- [1] 유현우, 조성배, “대화형 유전자 알고리즘을 이용한 감성기반 비디오 장면 검색,” *정보과학회논문지: 컴퓨팅의 실제*, 제10권, 제6호, 2004. 12.
- [2] W. Lin et al. (Eds.), *Multimedia Analysis, Processing and Communications*, Springer, 2011.
- [3] 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.
- [4] M. L. Smith, “Rapid Processing of Emotional Expressions without Conscious Awareness,” *Cerebral Cortex*, BHR250, 2011.
- [5] M. Feidakis, T. Daradoumis, S. Caballe, “Endowing e-Learning Systems with Emotion Awareness,” *Proc. of the 3rd International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, pp. 68-75 December 2011.
- [6] K. Holmqvist et al. (Eds.), *Eye tracking: A comprehensive guide to methods and measures*, Oxford, UK: Oxford University Press, 2011.
- [7] B. M. Herbert et al., “Cardiac awareness and autonomic cardiac reactivity during emotional picture viewing and mental stress,” *Psychophysiology*, Vol. 47, No. 2, pp. 342-354, 2010.
- [8] A. Bulling et al., “It’s in Your Eyes — Towards Context-Awareness and Mobile HCI Using Wearable EOG Goggles,” *Proc. of the 10th International Conference on Ubiquitous Computing (UbiComp 2008)*, pp. 84-93, ACM Press, 2008.
- [9] B. Blankertz, “Single-trial analysis and classification of ERP components — A tutorial,” *NeuroImage*, Vol. 56, No. 2, pp. 814-825, 2011.
- [10] J. M. Fernandes et al., “Video-EEG-fMRI: Contribution of in-bore Video for the Analysis of Motor Activation Paradigms,” *IFMBE Proceedings*, Vol. 25, No. 4, pp. 786-789, 2010.