

## 스토리 기반의 정보 검색 연구

유은순  
단국대학교 영화콘텐츠전문대학원  
(tesniere@naver.com)

박승보  
단국대학교 미디어콘텐츠연구원  
(molaal@naver.com)

.....

웹의 발전과 콘텐츠 산업의 팽창으로 비디오 데이터가 폭발적으로 증가함에 따라 데이터의 정보 검색은 매우 중요한 문제가 되었다. 그동안 비디오 데이터의 정보 검색과 브라우징을 위해 비디오의 프레임(frame)이나 샷(shot)으로부터 색채(color)와 질감(texture), 모양(shape)과 같은 시각적 특징(features)들을 추출하여 비디오의 내용을 표현하고 유사도를 측정하는 내용 기반(content-based)방식의 비디오 분석이 주를 이루었다. 영화는 하위 레벨의 시청각적 정보와 상위 레벨의 스토리 정보를 포함하고 있다. 저차원의 시각적 특징을 통해 내용을 표현하는 내용 기반 분석을 영화에 적용할 경우 내용 기반 분석과 인간이 인지하는 영화의 내용 사이에는 의미적 격차(semantic gap)가 발생한다. 왜냐하면 영화의 스토리는 시간의 진행에 따라 그 내용이 변하고, 관점에 따라 주관적 해석이 가능한 고차원의 의미정보이기 때문이다. 따라서 스토리 차원의 정보 검색을 위해서는 스토리를 모델링하는 정형화된 모형이 필요하다. 최근 들어 소셜 네트워크 개념을 활용한 스토리 기반의 비디오 분석 방법들이 등장하고 있다. 그러나 영화 속 등장인물들의 소셜 네트워크를 통해 스토리를 표현하는 이 방법들은 몇 가지 문제점들을 드러내고 있다. 첫째, 등장인물들의 관계에만 초점이 맞추어져 있으며, 스토리 진행에 따른 등장인물들의 관계 변화를 역동적으로 표현하지 못한다. 둘째, 등장인물의 정체성과 심리 상태를 보여주는 감정(emotion)과 같은 심층적 정보를 간과하고 있다. 셋째, 등장인물 이외에 스토리를 구성하는 사건과 배경에 대한 정보들을 반영하지 못하고 있다. 따라서 본 연구는 기존의 스토리 기반의 비디오 분석 방법들의 한계를 살펴보고, 문제 해결을 위해 문학 이론에서 제시하고 있는 서사 구조에 근거하여 스토리 모델링에 필요한 요소들을 인물, 배경, 사건의 세 가지 측면에서 제시하고자 한다.

.....

논문접수일 : 2013년 12월 05일    게재확정일 : 2013년 12월 19일  
투고유형 : 국문일반    교신저자 : 박승보

### 1. 서론

유튜브(youtube)와 블로그(blog)와 같이 사용자 제작 콘텐츠 서비스가 확산되면서 텍스트뿐만 아니라 멀티미디어 정보가 지속적으로 증가하고 있다. 여기에 전 세계적으로 문화콘텐츠 산업의 규모가 커짐에 따라 방대한 양의 영화와 애니메이션 등

이 제작, 소비되면서 비디오 데이터의 양적 성장이 가속화되고 있다. 조사에 의하면 2012년 전 세계에서 생성되는 디지털 정보의 양이 2.8제타 바이트에 이른 것으로 나타났다. 이처럼 기존의 일반적인 기술로 관리하기 힘든 빅데이터의 시대에서는 비디오 데이터의 검색과 축약의 성능을 향상시키기 위해 비디오의 내용을 분석하고 처리하는 연구에 대

\* 이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2013R1A1A2057943).

한 중요성이 더욱 커지고 있다. 과거에는 수작업으로 비디오의 내용적 의미를 키워드로 입력하는 주석 기반(annotation-based)의 방식을 활용하여 비디오 검색이 가능했지만(Lee et al. 2011) 비디오 데이터의 양이 폭발적으로 증가하면서 수작업으로 인한 고비용의 문제가 발생하고 많은 시간이 소요되는 한계가 발생하였다. 비디오의 프레임이나 샷(shot)으로부터 색채(color)와 질감(texture), 모양(shape)과 같은 시각적 특징을 추출하여 비디오의 내용을 표현하고 유사도를 측정하는 내용 기반(content-based)의 방식은 비디오 내용을 자동적으로 분석하고 처리함으로써 고비용과 시간 소요의 문제들을 해결하였다. 비디오 데이터는 시청각적 요소와 같은 저차원의 물리적 정보와 고차원의 의미적 정보를 포함하고 있다. 특히 영화나 애니메이션은 다른 비디오 데이터와는 달리 다층적 의미 정보인 스토리를 갖고 있다. 스토리는 시간의 진행에 따라 그 내용이 변화하는 특징이 있다. 또한 자동으로 처리하기 힘든 추상적 요소를 포함하고 있으며 주관적 해석이 가능한 가장 고차원의 의미 정보이기 때문에 스토리를 모델링하는데 많은 어려움이 따른다. 그동안 비디오 데이터 처리를 위해 시각적 특징과 같은 저차원의 물리적 정보를 이용한 내용 기반의 접근 방식이 주를 이루었다. 내용 기반 분석은 비디오의 계층적 구조를 다루기 위해 샷(shot)이나 장면(scene)의 경계를 검출하는 비디오 파싱에서부터 객체 인식, 이벤트 추출과 같이 비디오의 내용적 의미를 다루는 연구에 이르기까지 다양하게 진행되었다. 하지만 비디오의 시청각적 정보를 특징 벡터로 추출하여 비디오의 내용을 표현하는 내용 기반 분석은 영화와 같이 스토리 중심의 비디오 내용을 표현하는데 큰 문제를 드러낸다. 왜냐하면 비슷한 배경이나 오브젝트가 등장할 경우 해당 영화들은 비슷한 특징을 갖게 되기 때문에 스토리의 내용

과는 상관없이 유사한 영화로 판단하게 된다. 영화의 스토리는 시각적 특징만으로는 표현될 수 없다. 왜냐하면 스토리는 인과관계로 결합된 사건들의 집합인 플롯을 포함하고 있으며, 사건들을 일으키는 주체인 등장인물, 그리고 사건과 인물이 등장하는 물리적 공간 배경이 복합적으로 구성된 고차원적 의미 정보이기 때문이다. 이처럼 비디오의 내용을 시각적 요소들로 한정하여 표현하는 내용 기반 분석과 인간이 인지하고 이해하는 내용적 의미 사이에는 큰 차이가 발생하는데, 이것을 의미적 격차(semantic gap)(Hauptmann et al., 2007; Yang et al., 2007) 라고 한다. 시각적 특징의 단순함과 의미의 풍부함 사이에 존재하는 의미적 격차를 줄이기 위해 비디오에 등장하는 인물들의 얼굴 인식과 오브젝트 마이닝, 이벤트 추출과 같이 비디오의 내용을 의미적 차원에서 표현하려는 연구들이 시도 되었으나, 스포츠나 뉴스와 같은 일부 비디오에 적용 가능하며, 영화의 스토리를 표현하기에는 매우 미흡하다. 이를 해결하기 위해 스토리를 표현할 수 있는 정규화된 모형에 대한 연구가 필요한데, 최근 들어 소셜 네트워크 개념을 활용하여 비디오 속 등장인물들 간의 관계를 표현하는 새로운 유형의 스토리 모델링 방법들이 소개되고 있다. 소셜 네트워크 분석 방법을 영화의 등장인물 분석에 적용한 RoleNet(Weng et al., 2009)과 Character-net(Park, 2011; Park et al., 2012)은 영화 속 등장인물을 노드로 하고 그들 사이에 일어나는 대화나 출현을 간선으로 하는 소셜 네트워크를 생성함으로써 영화의 스토리를 모델링하였다. 이처럼 소셜 네트워크를 활용한 스토리 기반의 정보 검색 검색 방법은 기존의 내용 기반 방법과는 달리 스토리를 고려했다는 점에서 큰 의미가 있다. 그러나 RoleNet과 Character-net은 인물들 간에 누적된 관계를 정적으로 보여주는 것에 그치고 있으며 시간에

따른 인물들간의 관계 변화, 인물들의 내적 성향과 같은 스토리를 이해하는 데에 필요한 심층적인 정보를 표현하지 못하는 한계가 있다. 이에 본 연구는 기존의 스토리 기반의 비디오 정보 검색의 의의와 한계를 살펴보고, 문학에서 제시하고 있는 서사 구조를 토대로 스토리 모델링에 필요한 요소를 인물, 배경, 사건의 세 가지 측면에서 제시하고자 한다. 본 논문의 구성은 다음과 같다. 제 2장에서는 비디오 데이터의 정보 검색을 위한 기존의 내용기반 연구와 소셜 네트워크기반 연구 사례들을 살펴보고, 제 3장에서는 기존의 연구 방법들의 의의와 한계를 제시한다. 그리고 제 4장에서는 인문학적 서사 이론에 근거하여 스토리 모델링에 필요한 요소들을 인물, 배경, 사건의 세 가지 측면에서 제안한다.

## 2. 관련 연구

### 2.1 내용 기반 분석

비디오의 색상, 질감, 형태 등의 저차원의 시청각적 요소를 특징 벡터로 수치화하여 프레임이나 샷들 간의 유사도를 측정하는 내용기반 연구는 샷이나 장면의 경계를 검출하는 비디오 파싱에서부터 비디오가 포함하고 있는 객체들을 인식하고 주요 이벤트를 추출하는 의미 분석에 이르기까지 다양하게 진행되었다.

C. E. Nothelfer 등은 연도에 따른 영화의 샷의 길이 변화 비교를 위해 드라마와 코미디, 액션 장르의 영화들을 연도별로 12편을 선정하고 조명과 색채 정보를 이용하여 샷을 검출하였다. 연도에 따라 각 장르의 샷의 길이 변화를 살펴본 결과, 현대에 올수록 영화 속 샷의 길이가 짧아진다는 사실을 제시하였다(Nothelfer et al., 2009). F. Brodbeck은 영화의 샷으로부터 샷의 길이와 색상, 움직임 등을 추출하여 샷의 특징 값을 정한 후 샷들

을 원형으로 누적한 Fingerprint를 구축하였다(Brodbeck, 2011). Fingerprint를 통해 동일한 장르의 영화들을 비교하거나, 같은 감독이 연출한 영화들을 비교할 수 있다. 하지만 샷의 시각적 특징에만 의존했기 때문에 영화의 스토리를 구성하는 인물, 사건, 배경과 같은 의미적 내용 정보를 표현하지 못하는 한계가 있다. I. Laptev 등은 영화 속에서 배우들의 행위를 인식하는 외형 기반의 방법을 통해 배우들이 전화를 받거나 악수하기, 키스하기 등과 같은 8개의 이벤트들을 추출하였다(Laptev et al., 2008). 동영상에 등장하는 객체를 인식하고 객체들이 일으키는 이벤트를 통합한 연구 사례들도 있다. 박승보는 영화 속에 등장하는 인물들의 얼굴을 인식하고 얼굴의 입이 움직이는 인물을 화자로, 나머지 사람을 청자로 간주하여 두 인물들간에 이루어지는 대화 이벤트를 파악하였다(Park et al., 2009). V. Roth 역시 영화 속 인물들을 추출하고 각 인물들이 춤을 추는 장면, 폭파시키는 장면 등과 같은 행위 장면들을 추출하여 이를 시맨틱 네트워크로 표현하였다. 장면에 대한 시맨틱 네트워크를 통해 각 인물들간의 관계와 인물들의 행위들간의 관계들을 보여주고 있다(Roth, 1999).

시청각적 특징에만 의존한 내용 기반 비디오 연구는 영화와 애니메이션같이 스토리를 갖는 비디오의 내용적 의미를 충분히 표현할 수 없기 때문에 내용기반 연구와 비디오의 근본적인 의미 사이에는 항상 의미적 격차가 발생한다. 이러한 의미적 격차를 줄이기 위해 비디오 장르 분류와 비디오 내용 축약(abbreviation)과 같은 고차원적 의미 분석이 시도 되었다. Z. Rasheed 등은 샷의 길이와 색채 변화, 움직임과 조명과 같은 시각적 단서를 이용하여 영화를 4개의 장르(코미디, 액션, 드라마, 공포)로 분류하였다(Rasheed et al., 2005). 비디오 축약은 방대한 비디오 데이터로부터 원하

는 정보를 쉽게 찾을 수 있도록 불필요한 내용을 제거하고 데이터의 양을 축소하는 것이다(Otsuka et al., 2005; Wan et al., 2005). 비디오 축약은 비디오로부터 정지된 이미지인 키 프레임을 추출하여 표현하거나(Xie and Wu, 2008; Calic et al., 2007), 비디오 클립 형태의 비디오 스키밍(skimming)으로 표현한다(Ekin et al., 2003; Gong, 2003; Peker et al., 2006). 키프레임 기반의 비디오 요약은 타이밍이나 동기화 문제없이 내용을 빠르고 밀도 있게 나타낼 수 있는 장점이 있지만 샷의 대표 프레임을 시간적으로 나열하기 때문에 추격 장면과 같이 짧은 샷들이 연속될 때 짧거나 의미 없는 샷들의 키프레임도 일정량을 차지할 수 있는 문제가 있다. 비디오 스키밍은 오리지널 비디오에서 중요한 부분들을 추출하여 트레일러 혹은 비디오 클립과 같은 형태로 내용을 압축하는 것이다. A. Ekin 등은 축구 경기에서 골을 넣거나, 패널티 킥, 반칙 등의 중요한 이벤트를 중심으로 내용을 축약하였고(Ekin et al., 2003), K. A. Peker 등은 방송 비디오에서 얼굴 추적을 이용하여 내용을 요약하였다(Peker et al., 2006). Y. H. Gong는 시청각적 통합을 통한 비디오 요약을 시도했는데, 시각적 내용과 청각적 내용을 따로 분리한 후 이미지 트랙으로부터 뚜렷이 구별되는 장면을 추출하고, 오디오 트랙으로부터는 대표적인 문장을 추출한 후 이분그래프(bipartite graph)를 이용하여 추출한 시청각적 장면을 통합하였다(Gong, 2003). B. Jung 등은 영화나 드라마 같은 스토리 중심의 비디오 축약을 시도하였다. 대화, 액션과 같이 등장인물의 상호 작용이 나타나는 장면들과 편집적 관계에 의해 결정된 장면들을 추출한 후 장면을 노드로 하고 장면들 사이의 관계를 간선으로 하는 내러티브 구조 그래프를 구축하여 동영상 축약을 생성하였다(Jung et al., 2004).

## 2.2 소셜 네트워크 기반의 분석

비디오의 저차원적 요소인 시청각적 특징들을 추출하여 비디오의 내용을 표현하고 분석하는 기존의 내용기반 접근과 의미 기반 방식은 영화와 애니메이션 같이 스토리를 갖는 비디오에 대해 스토리의 의미를 고려하지 못하는 한계를 갖는다. 최근들어 고차원적 의미 정보인 스토리를 표현하기 위해 소셜 네트워크 개념을 활용하여 등장인물들 간의 관계를 분석하는 연구들이 활발히 진행되고 있다. 소셜 네트워크는 개인을 노드(node)로 하고 개인들 간의 관계를 간선(edge)으로 하는 네트워크 형태의 그래프이다(Wasserman and Faust, 1994; Son, 2002). 소셜 네트워크로부터 중요한 정보를 추출하기 위해 사용되는 다양한 개념들이 있다. 연결 정도(degree)는 한 노드가 주변의 다른 노드들과 얼마나 많은 관계를 갖고 있는지를 표현하는 개념으로 네트워크에서 중심에 위치하는 정보를 나타내는 중심성(centrality) 분석에 사용된다(Wasserman and Faust, 1994; Son, 2002). 인디그리(in-degree)는 어떤 노드가 다른 노드에게 화살표를 받는 정도를 의미하는 것으로 한 노드에게 수신되는 정보의 양을 표시한다. 아웃디그리(out-degree)는 어떤 노드가 다른 노드에게 화살표를 주는 정도를 의미하는 것으로 한 노드가 다른 노드들에게 정보를 제공하는 정도를 측정하는 지표이다(Wasserman and Faust, 1994; Son, 2002). 영화 속 스토리를 이끌어 나가는 등장인물은 대화와 행위를 통해 관계를 형성하며, 그 관계로부터 사건과 갈등이 발생한다. 영화는 등장인물들이 구축하는 관계를 통해 실제 세계의 사회적 구조를 재현한다. 따라서 소셜 네트워크 개념은 스토리 속 인물들 간의 관계와 인물의 중심성, 인물들 간에 발생하는 정보의 흐름 등을 분석하는데 매우 효과적이다.

### 2.2.1 등장인물의 소셜 네트워크 분석 연구

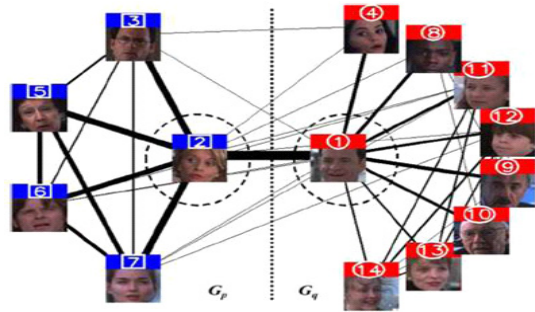
등장인물은 영화나 애니메이션처럼 스토리를 갖는 서사물에 등장하는 가공의 인물로써 사건, 배경과 함께 스토리를 구성하는 근본적인 요소이다.

R. Rienks 등은 회의에 참석한 참가자들의 말하는 행동과 상호작용, 역할 등과 같은 특성들로 구성된 사회적 관계를 이용하여 회의 결과에 가장 큰 영향을 미치는 영향력자를 가려내었고(Rienks et al. 2006), H. Hung 등은 회의 참가자가 발언한 시간과 동작 활성화도 같은 시청각적 특징을 이용하여 그룹 미팅에서 영향력자를 찾아내었다(Hung et al., 2007). 두 연구는 화자와 청자 사이에 일어나는 대화와 같은 상호작용을 이용하여 핵심 인물을 찾아냄으로써 소셜 네트워크로부터 유용한 정보를 추출할 수 있다는 것을 보여 주었다. 소셜 네트워크 개념들을 활용하여 스토리 속 주인공이나 주인공이 속한 커뮤니티와 같은 스토리 정보를 추출한 대표적인 연구 사례로는 RoleNet(Weng et al., 2009)과 Character-net(Park, 2011; Park et al., 2012)이 있다.

### 2.2.2 RoleNet

Weng 등은 얼굴 인식 기술을 이용하여 영화 장면에서 동시에 출현하는 등장인물들을 추출하고 그들 간의 연결을 누적하여 만든 RoleNet을 소개하였다. RoleNet은 등장인물이 노드가 되고, 등장인물들 간의 동시출현이 간선이 되며, 동시 출현 장면의 수가 가중치로 표현되는 소셜 네트워크이다(Weng et al., 2009). 이렇게 구축된 RoleNet으로부터 인물들의 비중도와 같은 중요한 정보를 추출하기 위해 중심성 분석을 이용하여 주연과 조연을 분류하였다. 주연과 조연을 구분하는 경계 값은 노드의 연결정도 중심성 값을 계산하여 정렬한 후

최대 간극을 나타내는 노드에 의해 결정된다. 또한 Max-Flow Min-Cut 알고리즘을 사용하여 주연과 함께 주연이 속한 커뮤니티도 추출하였다. <Figure 1>은 영화 ‘You’ve got mail’의 등장인물들 간의 관계를 RoleNet으로 표현한 것인데, 2명의 남녀 주연과 각각의 커뮤니티를 보여주고 있다. RoleNet은 가중치의 단위를 등장인물이 출현하는 장면의 수로 설정하였다. 하지만 장면에 등장하는 대화의 양이 일정하지 않고, 장면의 진행 시간 또한 크게 차이가 난다. 즉 장면에 포함된 정보의 양은 장면에 따라 다르기 때문에 장면이 아닌 대화의 양을 관계의 기본 단위로 설정하는 것이 더 적합하다.

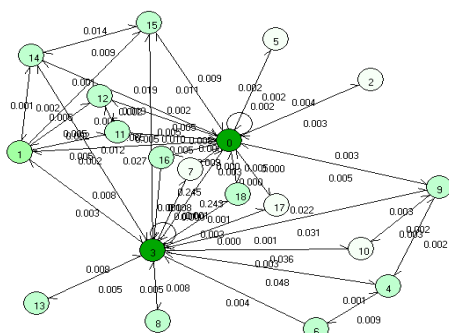


<Figure 1> Major Characters and Community from RoleNet of Movie ‘You’ve got mail’

### 2.2.3 Character-net

Park 등은 인물들의 관계를 장면 대신 대화로 설정하고 대본이나 자막을 이용하여 영화 속 등장인물들 간의 대화를 누적인 Character-net을 제시하였다. Character-net은 영화의 등장인물을 노드로 하고 등장인물 사이의 대화 방향과 대화량을 각각 간선과 가중치로 표현한 소셜 네트워크이다(Park, 2011; Park et al., 2012). 등장인물의 동시출현과 동시출현 장면의 개수를 간선과 가중치로

사용했던 RoleNet과 달리 Character-net은 대화를 간선과 가중치로 사용했기 때문에 화자와 청자라는 대화의 방향성이 존재하며 정보의 흐름을 추출할 수 있다. <Figure 2>는 영화 'Pretty Woman'의 등장인물들을 Character-net으로 표현한 것이다. Character-net으로부터 정보를 추출하기 위해 중심성 분석을 이용하여 등장인물들을 주연과 조연, 단역으로 분류하였다. 비중도에 따른 등장인물들의 분류는 전체 노드들의 중심성 평균값을 구한 후 평균 이하의 노드들은 단역으로, 평균 이상의 노드들은 주연 또는 조연으로 설정하였다. 그리고 주연과 조연을 구분하기 위해 평균 이상의 노드들을 중심성 값에 의해 내림차순으로 정렬한 후 최대 간극을 갖는 노드들을 찾아 간극 이상의 노드들은 주연으로, 이하의 노드들은 조연으로 분류하였다(Park, 2011; Park et al., 2012).



<Figure 2> Character-net of Movie 'Pretty Woman'

RoleNet과 Character-net 이외에도 L. Ding과 A. Yilmaz는 영화 대본을 이용하여 등장인물들 간의 소셜 네트워크를 구축하고 중심성 분석과 연결 정도 분석을 통해 등장인물들의 커뮤니티와 각 커뮤니티의 리더를 찾아내었다(Ding and Yilmaz, 2010). J. Kaminski과 M. Schober는 1915년부터 2011년까지의 영화 797편을 선별하여 영화 대본

으로부터 등장인물 이름과 대사를 이용하여 소셜 네트워크를 생성하였다. 영화의 장르와 제작시기 등과 같은 변수들에 따라 영화의 소셜 네트워크 구조들 간에 차이가 있다는 것을 시각적으로 제시하였다(Kaminski and Schober 2011). 소셜 네트워크 기반의 스토리 모델들은 비디오의 내용적 의미를 표현한 고차원적 접근이라는 점에서 의의가 있으며, 특히 RoleNet과 Character-Net은 중심성 분석을 통해 등장인물의 중요성을 판별하는 방법론을 제시하여 등장인물이 스토리에서 차지하는 비중을 수치로 표시하였다는 점에서 의미가 있다. 이처럼 등장인물의 소셜 네트워크를 통한 스토리 모델링은 등장인물의 비중도와 같은 스토리의 중요한 정보를 추출할 수 있지만 이는 등장인물들간의 누적된 결과를 보여줄 뿐 스토리가 진행되는 동안 일어나는 등장인물들의 출현과 그들의 관계 변화를 보여주지 못한다는 한계가 있다. 스토리를 구성하는 요소인 공간적 배경에 대한 정보 또한 누락되어 있다.

### 3. 기존 연구들에 대한 문제 제기

제 2장에서는 비디오 데이터에 대한 전통적인 내용기반 연구에서부터 스토리를 고려한 최근의 소셜 네트워크 기반의 연구에 이르기까지 비디오 데이터를 모델링하기 위한 연구 사례를 검토하였다. 3장에서는 기존의 연구 방법에 대한 문제점을 구체적으로 살펴보고, 스토리를 모델링하기 위한 방향을 제시하고자 한다. 초기의 내용기반 연구의 주요 문제점들은 다음과 같다. 첫째, 내용(content)에 대한 정의와 표현 방법의 문제이다. 비디오는 물리적 정보인 시청각적 요소와 스토리 같은 의미적 정보를 포함하고 있다. 특히 스토리는 시간의 진행에 따라 변화하며, 관점에 따라 주관적

해석이 가능한 고차원의 의미 정보이기 때문에 스토리를 모델링하는데 많은 어려움이 따른다. 내용 기반 연구는 비디오의 내용을 정의하고 그것을 표현하는 방식에 있어서 중요한 문제를 드러낸다. 내용 기반 연구는 프레임이나 샷에 포함된 오브젝트나 배경의 색채와 질감, 모양과 같은 저차원의 시각적 특징을 이용하여 내용을 표현하는데, 만약 시각적 특징만을 내용으로 정의하게 되면 비슷한 배경과 오브젝트들이 등장하는 영화들의 경우 칼라 히스토그램이 유사하게 나오기 때문에 스토리의 내용과는 상관없이 유사한 영화로 판단될 수 있다. 영화의 스토리는 단순히 시각적 특징으로만 설명될 수 없는 여러 가지 요소들이 복합적으로 구성되어 있다. 또한 스토리는 슬픔, 욕망과 같이 특징 벡터로 표현하기 어려운 추상적인 내용들도 포함하고 있다. 내용기반 연구는 비디오의 내용을 색채와 같은 물리적인 특징들로 한정하여 표현하기 때문에 스토리 차원의 의미를 간과하는 문제를 갖는다. 둘째, 의미적 격차의 문제를 들 수 있다. 인간이 비디오의 내용을 인지하고 이해하는 과정은 단계적으로 이루어지는데, 먼저 대략적이고 심플한 의미적 객체들을 감지하고 그것들을 조합함으로써 내용의 의미를 파악하는 과정으로 이루어진다(Yang et al., 2007). 따라서 시각적 특징들을 통해 내용을 표현하는 내용 기반 연구와 인간에 의해 인지된 비디오의 내용 사이에는 의미적 격차가 발생한다. 시청각적 요소의 단순함과 의미의 다양함 사이에 존재하는 의미적 격차를 줄이기 위해 객체 인식과 오브젝트 마이닝과 같은 의미적 정보를 다루는 연구들이 진행되었으나 스포츠나 뉴스와 같은 특정 비디오 분석에 적용 가능하며, 스토리를 갖는 영화를 분석하기에는 한계를 보인다. 의미적 격차를 줄이기 위해서는 스토리를 표현할 수 있는 정규화된 모형에 대한 연구가 필요

하다. 스토리 모델링을 위해 최근에 소셜 네트워크 개념을 활용한 등장인물 간 관계를 분석하는 연구들이 활발히 진행되었다. 현재 소셜 네트워크를 활용하여 영화 속 등장인물의 관계와 등장인물이 속한 커뮤니티를 표현하는 연구들은 의미적 단위인 스토리를 고려했다는 점에서 기존의 연구 방법들과 분명한 차별화를 이루었다. 그러나 스토리가 갖고 있는 주관적이고 추상적인 속성, 그리고 인과 관계로 연결되어 있는 스토리의 사건들을 표현하기에는 다음과 같은 한계들을 갖는다. 첫째, 현재 소개되고 있는 RoleNet이나 Character-net은 등장인물들의 관계에 대한 누적된 결과를 보여주는 정적인 표현 방법이다. 스토리는 시간의 진행에 따라 변화하기 때문에 스토리 변화에 따라 형성되는 등장인물들의 역동적인 관계 변화를 보여줘야 한다. 둘째, 중심성과 밀도와 같은 개념을 이용하여 등장인물들의 소셜 네트워크로부터 주연과 조연, 단역을 추출하여 등장인물들을 비중도에 따라 분류할 수 있지만 스토리가 전개되는 동안 등장인물이 겪게 되는 정서적 변화와 정신적, 물리적 성장과 같은 심층적인 의미 정보는 표현하지 못한다. 셋째, 공간적 배경 정보의 누락이다. 배경은 인물, 사건과 함께 스토리를 구성하는 주요 요소이기 때문에 등장인물들의 만남과 사건이 일어나는 공간적 정보를 고려해야 한다.

#### 4. 서사구조에 근거한 스토리 모델링 요소

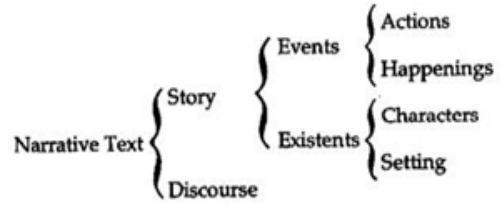
스토리 차원에서 영화들 간의 유사도를 비교하거나 영화로부터 스토리 정보를 검색하기 위해서는 정형화된 스토리 모델이 필요하다(Park et al., 2011). 본 연구에서는 영화와 같이 스토리 중심의 비디오 데이터에 대한 기존의 연구에서 간과하고

있는 스토리 모델링에 필요한 주요 요소들을 제안하고자 한다.

#### 4.1 서사(Narrative)의 구조

스토리의 구성 요소들을 이해하고 형식화하기 위한 노력들은 매우 오래전부터 진행되어 왔다. 아리스토텔레스(Aristoteles)는 <시학>에서 스토리를 3막 구조인 전반부-중반부-후반부로 나누었고, 러시아 형식주의자 Vladimir Propp은 민담 속 등장인물들이 행하는 31개의 전형적인 기능을 추출하였다. 또한 신화학자 Joseph Campbell은 세계 각 문화권의 영웅 신화를 분석하여 모든 영웅 신화 속 영웅들이 공통적으로 경험하는 영웅의 모험 스키마(Hero's Journey schema)를 추출하였다. 많은 학자들은 서사 구조를 이분법적 측면에서 다루었다. 대표적인 예로 Seymour Chatman은 서사를 일련의 사건들과 등장인물, 배경을 포함하는 '스토리'와 이야기를 표현하는 형식인 '담화(discourse)'로 분리하였다. 러시아 형식주의자들은 '이야기'와 '담화'라는 용어 대신 각각 '파블라(fablula)'와 '쉬젯트(sjuzet)'라는 용어를 사용하였고, 프랑스 구조주의자들은 '이스뚜아르(histoire)'와 '디스꾸르(discours)', 그리고 영국의 평론가 E. M. Forster는 '스토리'와 '플롯(plot)'이라는 용어를 사용하며 서사의 이원성을 설명하였다.

<Figure 3>은 챗먼이 제시한 서사의 이원적 구조이다. 서사는 이야기(story)와 담화(discourse)로 구성되어 있으며, 이야기(story)는 다시 사건들(events)과 존재물들(existents)로 나뉜다. 사건들은 인물들의 행위(actions)와 사고(happening)로 구성되며, 존재물들은 인물(character)과 인물들이 위치하는 시공간적 배경(setting)으로 이루어진다(Chatman, 1990).

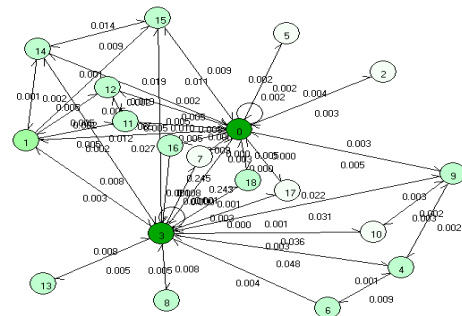


<Figure 3> Chatman's Dual Structure of Narrative

본 연구는 위에서 제시한 챗먼의 서사 구조에 근거하여 스토리 모델링에 필요한 주요 요소들을 등장인물과 사건, 그리고 배경의 세 가지 측면에서 제시하고자 한다.

#### 4.2 등장인물 측면 : 등장인물의 감정어 추출을 통한 감정 패턴 모델링

<Figure 4>는 영화 'Pretty Woman'의 등장인물들의 대화(dialogue)를 누적하여 그들간의 관계를 소셜 네트워크로 표현한 Character-net이다. 중심성과 밀도 분석을 통해 영화의 남녀 주인공인 '에드워드(0번)'와 '비비안(3번)'을 추출하였다. Character-net은 등장인물들의 누적된 결과를 통해 등장인물들의 복잡하고 다층적인 관계를 쉽게 이해할 수 있도록 직관적으로 전달하고 있지만 등장인물의 심리적 상황이나 내적 성격과 같이 스토리를 이해하는데 필요한 심층적인 정보는 표현하지

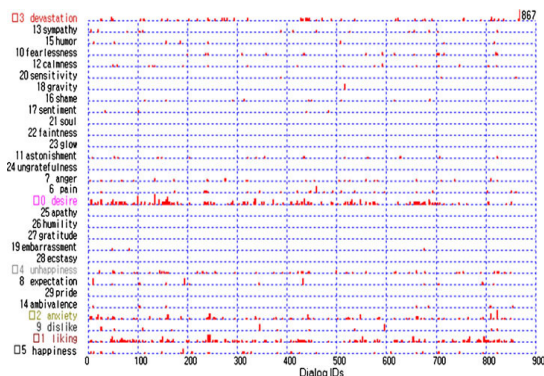


<Figure 4> Character-net of Movie 'Pretty Woman'



못하는 한계를 갖는다.

이 문제를 해결하기 위해 영화의 대사로부터 등장 인물의 감정을 추출하여 인물들의 근본적인 성격이나 감정의 패턴 양상을 모델링하는 것이 필요하다. 영화의 대사는 등장인물들의 감정을 전달하는 중요한 도구이며, 대사를 통해 발현된 감정어는 인물들의 정체성과 심리 상태를 직접적으로 드러내 준다. 따라서 감정의 흐름은 등장인물과 그의 동기를 이해하는데 유용한 정보이다. 추출한 감정어를 긍정과 부정의 두 개의 극성으로 분류하거나, 좀 더 세분화된 분류를 위해 심리학이나 철학에서 제시된 다양한 감정 범주를 이용하여 분류한다. <Figure 5>는 대규모 영어 어휘 데이터베이스인 워드넷(WordNet)을 이용하여 영화 'Pretty Woman'의 대본(script)으로부터 인물들의 감정어들을 자동 추출하여 30개의 감정 범주로 분류한 것이다. 감정어 분석을 통해 스토리 진행에 따라 변화하는 등장 인물들의 감정 패턴 양상을 시각화해줄 수 있다.



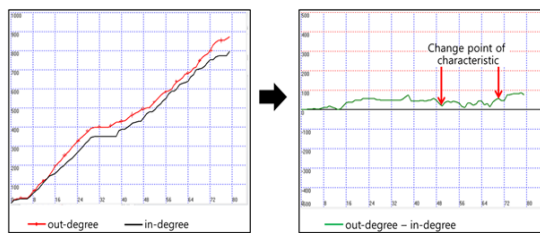
<Figure 5> Emotion Words Extraction of Movie 'Pretty Woman'

#### 4.3 사건의 측면 : 캐릭터 아크(Character Arc) 모델링을 위한 성향 분석

스토리는 인과관계로 연결된 일련의 사건들이며,

사건들의 일관성 있고 논리적인 배열을 플롯이라고 한다. 사건은 갈등을 바탕으로 하며 등장인물들에 의해 일어나고 그들에게 영향을 미침으로써 그들의 정체성과 입체적인 성격을 구체적으로 규정한다. 기존의 내용기반 연구에서는 시각적 특징과 등장인물의 액션과 어휘를 활용하여 주요 이벤트들을 추출하였다. 하지만 사건이 주인공에게 어떻게 영향을 미치고 변화를 야기했는지에 대한 정보는 표현하지 못했다. 로버트 맥기(Robert Mckee)에 의하면 사건은 등장인물들의 삶의 조건에 의미 있는 변화를 일으키며 이 변화는 가치의 변화라는 형태로 경험되고 표현된다고 하였다(Mckee, 2002). 예를 들어 로맨틱 코미디 속 수동적이었던 주인공이 적극적인 여성으로 변화하면서 사랑을 얻고, 액션 히어로 영화 속 소심한 주인공이 일련의 사건을 겪으면서 악과 맞서는 영웅의 모습으로 변화한다. 영화 'Pretty Woman'의 남자 주인공 에드워드 역시 영화 초반에는 기업의 이익만을 생각하는 차갑고 이기적인 사업가로 등장하지만 주인공 비비안을 만나면서 영화 후반부에는 따뜻하고 남을 배려하는 이해심 많은 인물로 변화한다. 이처럼 스토리가 전개되면서 등장인물이 정서적, 심리적, 정신적으로 변화를 겪으면서 성장하고 발전하는 것을 캐릭터 아크 혹은 변형적 아크(transformational arc)(Cowgill, 2003; Marks, 2007)라고 부른다. 캐릭터 아크는 영화에서 큰 의미를 가진다. 왜냐하면 사건은 등장인물이 갈등을 해결해 나가는 과정에서 어떻게 변화하고 발전하는지를 보여주며, 관객은 그들의 성장과 발전을 통해 영화의 의미를 이해하게 되기 때문이다. 캐릭터 아크를 모델링하기 위해서는 사건에 따른 주인공들의 내적 성향 분석이 선행되어야 한다. 등장인물의 성향 변화 여부는 소셜 네트워크 개념을 이용하여 분석할 수 있다. 소셜 네트워크의 인디그리와 아웃디그리와 같은 개념을 이용하여 인물

들의 중심성 변화를 추출한다. 이 개념을 영화 속 등장인물 분석에 적용하면, 인디그리는 한 인물이 다른 인물들로부터 말을 듣는 것을 의미하고, 아웃디그리는 한 인물이 다른 인물들에게 말을 하는 것을 의미한다. 따라서 인디그리가 높으면 다른 인물들로부터 말을 많이 듣는 것이고 반대로 아웃디그리가 높으면 다른 인물들에게 말을 많이 하는 것을 의미한다. 따라서 인디그리와 아웃디그리의 시계열적 변화를 통해 주인공의 변화 양상을 추출한다. <Figure 6>은 영화 'Pretty Woman'의 남자 주인공 에드워드의 대사를 분석하여 인디그리와 아웃디그리의 변화를 표현한 것이다.



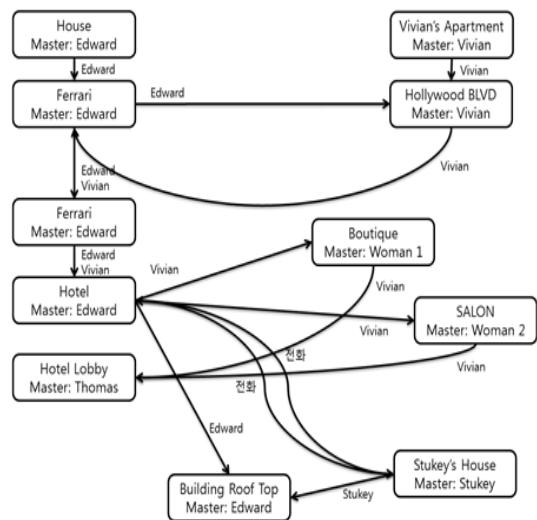
<Figure 6> Change Point of Characteristic for Movie Character 'Edward'

우측의 그림은 아웃디그리에서 인디그리를 뺀 그래프로 정보의 제공 및 수신 성향을 나타내는 그래프이다. 영화 전반부에서 정보를 제공하는 패턴에서 중후반부에서 정보를 수신하는 패턴으로 변화하는 부분에서 에드워드의 성향이 변화하는 것으로 판단할 수 있다. 이렇게 추출된 성향 변곡점과 아웃디그리와 인디그리 사이의 차이를 이용하여 등장인물의 변화와 성장을 표현하는 캐릭터 아크를 추출할 수 있는 연구를 진행할 필요가 있다.

#### 4.4 배경적 측면 : 공간의 배치 및 거리

배경은 이야기가 진행되는 물리적 공간이다. 따

라서 등장인물들의 만남과 대화가 이루어지고 사건이 벌어지는 공간적 배경은 스토리의 내용적 의미를 구성하는 중요한 정보이다. 영화 'Pretty Woman'에서 남녀 주인공의 첫 만남이 이루어지는 거리와 호텔, 서로를 알아가면서 이해하는 과정이 담긴 오페라 극장과 공원, 그리고 사랑을 확인하는 여주인공의 아파트 등의 공간들은 영화에서 중요한 의미를 담고 있는 공간이다. 이처럼 영화 속 의미적 공간 정보를 추출하기 위해 영화의 대본을 활용할 수 있다. 영화 대본의 세계적인 표준으로 사용되고 있는 Final Draft는 장면의 시작을 표시하고 촬영된 장소와 시간을 지시하는 'scene heading'과 등장인물의 소개와 행위를 설명하는 'action', 등장인물의 이름을 표시하는 'character name', 그리고 등장인물의 대사를 의미하는 'dialogue' 등으로 구성되어 있다. 등장인물들의 단독 등장 혹은 다른 인물들과의 동시 출현에 따라 공간적 배경을 클러스터링하고 시계열적으로 배치함으로써 <Figure 7>과 같이 스토리의 공간 맵을 모델링한다.



<Figure 7> Scene Map of Movie 'Pretty Woman'

## 5. 결론

본 연구는 비디오의 저차원적 정보를 통해 고차원의 내용적 의미를 표현하려는 내용 기반의 비디오 분석에서부터 최근에 등장한 스토리를 고려한 소셜 네트워크 기반의 비디오 분석 방법에 이르기까지 비디오 데이터 처리를 위해 그동안 진행되어 온 다양한 연구 사례를 통해 각 연구 방법들이 갖는 한계를 살펴보았다. 기존의 내용 기반 방식이 갖는 가장 큰 문제점은 비디오의 내용을 시각적인 특징들로 간주하기 때문에 영화와 같이 스토리 중심의 비디오에 적용할 경우 스토리의 내용적 의미를 간과하는 점이다. 이처럼 스토리를 고려하지 못하기 때문에 실제 영화의 내용과 추출된 정보 사이에는 의미적 격차가 발생하게 된다. 의미적 격차를 줄이기 위해 최근에 등장한 소셜 네트워크 기반의 접근 방식은 등장인물들 간의 소셜 네트워크를 생성하여 스토리를 표현한다는 점에서 기존의 접근 방식과는 다른 스토리 차원의 접근 방식이다. 하지만 스토리는 시간의 진행에 따라 그 내용이 변하는 속성이 있으며 스토리를 보는 관점에 따라 주관적 해석이 가능하기 때문에 스토리를 모델링하는데 많은 어려움이 따른다. 본 논문은 영화로부터 스토리 정보를 검색하고 스토리 차원에서 영화들 간의 유사도를 비교하기 위해 정형화된 스토리 모델 구축에 필요한 요소들을 서사의 구성 요소에 근거하여 인물, 배경, 사건의 세 가지 측면에서 제안하였다. 앞으로 등장인물들 간의 시계열적 변화 패턴이나 스토리의 변화 지점을 자동적으로 검색할 수 있는 비디오 검색 시스템 설계를 위해서는 본 연구에서 제시한 등장인물의 감정 패턴과 캐릭터 아크, 공간적 정보 이외에도 스토리에 대한 심층적인 분석과 이해가 필요하며, 좀 더 거시적인 관점에서 서사와 컴퓨터 공간간의 융합적인 연구가 이루어져야 한다.

## 참고문헌

- Brodbeck, F., *Cinematics*, 2011, Available at <http://cinematics.fredericbrodbeck.de> (Downloaded 30 September, 2013).
- Calic, J., D. Gibson and N. Campbell, "Efficient layout of comic-like video summaries," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.17, No.7(2007), 931~936.
- Chatman, S., *Story and Discourse: Narrative Structure in Fiction and Film*, Minumsa, Seoul, 1990.
- Cowgill, L., *Secrets of Screenplay Structure*, Simgongart, Seoul, 2003.
- Ding, L. and A. Yilmaz, "Learning Relations Among Movie Characters : A Social Network Perspective," *Proceedings of the 11th European conference on Computer vision : Part IV (ECCV)*, (2010), 410~423.
- Ekin, A., A. M. Tekalp and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Transactions on Image Processing*, Vol.12, No.7(2003), 796~807.
- Gong, Y. H., "Summarizing audio-visual contents of a video program," *EURASIP Journal on Advances in Signal Processing*, Vol.2003, No.2(2003), 160~169.
- Hauptmann, A., R. Yan and W-H. Lin, "How many high-level concepts will fill the semantic gap in video retrieval?," *Proceedings of the 6th ACM international Conference on Image and Video Retrieval*, (2007), 627~634.
- Hung, H., D. Jayagopi, C. Yeo, G. Friendland, S. Ba, J. Ramchandran, N. Mirghafori and D. Gatica-Perez, "Using audio and video features to classify the most dominant person in a group meeting," *Proceedings of ACM Multimedia Conference*, (2007), 835~838.

- Jung, B., T. Kwak, J. Song and Y. Lee, "Narrative abstraction model for story-oriented video," *Proceedings of ACM Multimedia Conference*, (2004), 828~835.
- Kaminski, J. and M. Schober, "Social networks in movies," *International Conference on Collaborative Innovation Networks COINs*, (2011), 1~3.
- Laptev, I., M. Marszałek, C. Schmid and B. Rozenfeld, "Learning realistic human actions from movies," *IEEE Conference on Computer Vision and Pattern Recognition*, (2008), 1~8.
- Lee, Y.-H., K.-J. Oh, S. Visal, and G.-S. Jo, "A Collaborative Video Annotation and Browsing System using Linked Data," *Korea Intelligent Information System Society*, Vol.17, No.3(2011), 203~219.
- Marks, D., *Inside Story*, Three Mountain Press, 2007.
- Mckee, R., *Story : Substance, Structure, Style and the Principles of Screenwriting*, Golden Bough, Seoul, 2002.
- Nothelfer, C. E., J. E. DeLong and J. E. Cutting, "Shot Structure in Hollywood Film," *Indiana Undergraduate Journal of Cognitive Science*, Vol.4(2009), 103~113.
- Otsuka, I., K. Nakane, A. Divakaran, K. Hatanaka and M. Ogawa, "A highlight scene detection and video summarization system using audio feature for a personal video recorder," *IEEE Trans. Consumer Electronics*, Vol.51, No.1, (2005), 112~116.
- Park, S.-B., *Semantic Multimedia Browsing System based on Character-net*, PhD Thesis, Department of Information Engineering, INHA University, 2011.
- Park, S.-B., E.-S. You and J. Jung, "Extracting Beginning Boundaries for Efficient Management of Movie Storytelling," *Journal of Intelligence and Information Systems*, Vol.17, No.4(2011), 279~292.
- Park, S.-B., K.-J. Oh and G.-S. Jo, "Social Network Analysis in a Movie using Character-net," *Multimedia Tools and Applications*, Vol.59, No.2(2012), 601~627.
- Park, S.-B., Y.-W. Kim and G.-S. Jo, "Conversation Context Annotation using Speaker Detection," *Korea Multimedia Society*, Vol.12, No.9(2009), 1252~1261.
- Peker, K. A., I. Otsuka and A. Divakaran, "Broadcast video program summarization using face tracks," *IEEE International Conference on Multimedia and Expo*, (2006), 1053~1056.
- Rasheed, Z., Y. Sheikh and M. Shah, "On the use of computable features for film classification," *Circuits and Systems for Video Technology, IEEE Transactions on*, Vol.15, No.1 (2005), 52~64.
- Rienks, R., D. Zhang and W. Post, "Detection and application of influence rankings in small group meetings," *Proceedings of International Conference on Multimodal Interfaces*, (2006), 257~264.
- Roth, V., "Content-based retrieval from digital video," *Image and Vision Computing*, Vol.17 (1999), 531~540.
- Son, D. W., *Social Network Analysis*, Kyungmunsa, Seoul, 2002.
- Wan, K., X. Yan and C. Xu, "Automatic mobile sports highlights," *IEEE International Conference on Multimedia and Expo(ICME)*, (2005), 638~641.
- Wasserman, S. and K. Faust, *Social Network Analysis : Methods and Applications*, Cambridge University Press, 1994.
- Weng, C. Y., W. T. Chu and J. L. Wu, "RoleNet:

- movie analysis from the perspective of social network," *IEEE Transaction on Multimedia*, Vol.11, No.2(2009), 256~271.
- Xie X.-N. and F. Wu, "Automatic video summarization by affinity propagation clustering and semantic content mining," *2008 International Symposium on Electronic Commerce and Security*, (2008), 203~208.
- Yang, S. G., S. K. Kim and Y. M. Ro, "Semantic home photo categorization," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.17, No.3(2007), 324~335.

Abstract

## Story-based Information Retrieval

Eun-Soon You\* · Seung-Bo Park\*\*

Video information retrieval has become a very important issue because of the explosive increase in video data from Web content development. Meanwhile, content-based video analysis using visual features has been the main source for video information retrieval and browsing. Content in video can be represented with content-based analysis techniques, which can extract various features from audio-visual data such as frames, shots, colors, texture, or shape. Moreover, similarity between videos can be measured through content-based analysis. However, a movie that is one of typical types of video data is organized by story as well as audio-visual data. This causes a semantic gap between significant information recognized by people and information resulting from content-based analysis, when content-based video analysis using only audio-visual data of low level is applied to information retrieval of movie. The reason for this semantic gap is that the story line for a movie is high level information, with relationships in the content that changes as the movie progresses. Information retrieval related to the story line of a movie cannot be executed by only content-based analysis techniques. A formal model is needed, which can determine relationships among movie contents, or track meaning changes, in order to accurately retrieve the story information. Recently, story-based video analysis techniques have emerged using a social network concept for story information retrieval. These approaches represent a story by using the relationships between characters in a movie, but these approaches have problems. First, they do not express dynamic changes in relationships between characters according to story development. Second, they miss profound information, such as emotions indicating the identities and psychological states of the characters. Emotion is essential to understanding a character's motivation, conflict, and resolution. Third, they do not take account of events and background that contribute to the story. As a result, this paper reviews the importance and weaknesses of previous video analysis methods ranging from content-based approaches to story analysis based on social network. Also, we suggest necessary elements, such as character, background, and events, based on narrative structures introduced in the literature. We extract characters' emotional words from the script of the movie *Pretty*

---

\* Graduate School of Cinematic Content, Dankook University

\*\* Corresponding Author: Seung-Bo Park

Institute of Media Content, Dankook University

126 Jukjeon-dong, Suji-gu, Yongin-si, Gyeonggi-do, 448-701, Korea

Tel: +82-31-8005-2387, Fax: +82-31-8021-7249, E-mail: molaal@naver.com

Woman by using the hierarchical attribute of WordNet, which is an extensive English thesaurus. WordNet offers relationships between words (e.g., synonyms, hypernyms, hyponyms, antonyms). We present a method to visualize the emotional pattern of a character over time. Second, a character's inner nature must be predetermined in order to model a character arc that can depict the character's growth and development. To this end, we analyze the amount of the character's dialogue in the script and track the character's inner nature using social network concepts, such as in-degree (incoming links) and out-degree (outgoing links). Additionally, we propose a method that can track a character's inner nature by tracing indices such as degree, in-degree, and out-degree of the character network in a movie through its progression. Finally, the spatial background where characters meet and where events take place is an important element in the story. We take advantage of the movie script to extracting significant spatial background and suggest a scene map describing spatial arrangements and distances in the movie. Important places where main characters first meet or where they stay during long periods of time can be extracted through this scene map. In view of the aforementioned three elements (character, event, background), we extract a variety of information related to the story and evaluate the performance of the proposed method. We can track story information extracted over time and detect a change in the character's emotion or inner nature, spatial movement, and conflicts and resolutions in the story.

**Key Words** : Video Data, Content-Based Analysis, Semantic Gap, Story, Social Network

## 저 자 소개



**유은순**

인하대학교 불어불문학과를 졸업하였고, 2001년과 2007년에 프랑스 Franche-Comte 대학교에서 언어학(세부전공 : 자연언어처리)석사 학위와 박사 학위를 각각 취득하였다. 2011년부터 2012년 8월까지 단국대학교 미디어콘텐츠연구원 스토리텔링 연구센터의 전임연구원으로 근무하였고, 2012년 9월부터 현재까지 단국대학교 영화콘텐츠전문대학원 리서치 펠로우로 재직 중이다. 연구 관심분야는 시맨틱 웹, 온톨로지, 디지털 스토리텔링, 소셜 네트워크, 빅데이터 등이다.



**박승보**

인하대학교 전기공학과를 1995년에 졸업하였고, 1997년에 인하대학교 전기공학과에서 석사학위를, 2011년에 인하대학교 정보공학과에서 박사 학위를 각각 취득하였다. 1996년부터 2002년까지 대우전자에서 주임연구원으로 근무하였고, 2011년부터 2012년까지 인하대학교 교육대학원에서 강사로 재직하였다. 2013년 7월부터 현재까지 단국대학교 미디어콘텐츠연구원 연구원으로 재직 중이다. 연구 관심분야는 비디오 스토리 분석, 스토리텔링 분석, 시맨틱 콘텐츠, 비디오 지식표현, 소셜 네트워크 분석, 인공지능 등이다.