지능정보연구 제18권 제2호 2012년 6월(pp.47~60)

시각적 특징을 기반한 샷 클러스터링을 통한 비디오 씬 탐지 기법

신동욱
한양대학교 컴퓨터공학과
(foremostdw@gmail.com)

김태환
한양대학교 BK21 AIS 사업단
(kimth4110@gmail.com)

최중민
한양대학교 컴퓨터공학과
(jmchoi@hanyang.ac.kr)

비디오 데이터는 구조화되지 않은 복합 데이터의 형태를 지닌다. 이러한 비디오 데이터의 효율적인 관리 및 검색을 위한 비디오 데이터 구조화의 중요성이 대두되면서 콘텐츠 내 시각적 특징을 기반으로 비디오 씬(scene)을 탐지하고자 하는 연구가 활발히 진행되었다. 기존의 연구들은 주로 색상 정보만을 이용하여 샷(shot) 간의 유사도 평가를 기반한 클러스터링(clustering)을 통해 비디오 씬을 탐지하고자 하였다. 하지만 비디오 데이터의 색상 정보는 노이즈(noise)를 포함하고, 특정 사물의 개입 등으로 인해 급격하게 변화하기 때문에 색상만을 특징으로 고려할 경우, 비디오 샷 혹은 씬에 대한 올바른 식별과 디졸브(dissolve), 페이드(fade), 와이프(wipe)와 같은 화면의 점진적인 전환(gradual transitions) 탐지는 어렵다. 이러한 문제점을 해결하기 위해, 본 논문에서는 프레임(frame)의 컬러 히스토그램과 코너 에지, 그리고 객체 컬러 히스토그램에 해당하는 시각적 특징을 기반으로 동일한 이벤트를 구성하는 의미적으로 유사한 샷의 클러스터링을 통해 비디오 씬을 탐지하는 방법(Scene Detector by using Color histogram, corner Edge and Object color histogram, SDCEO)을 제안한다.

SDCEO는 샷 바운더리 식별을 위해 컬러 히스토그램 분석 단계에서 각 프레임의 컬러 히스토그램 정보를 이용하여 1차적으로 연관성 있는 연속된 프레임을 샷 바운더리로 병합한 후, 코너 에지 분석 단계에서 병합된 샷 내 처음과 마지막 프레임의 코너 에지 특징 비교를 통하여 샷 바운더리를 정제하여 최종 샷을 식별한다. 키프레임 추출 단계에서는 샷 내 프레임간 유사도 비교를 통해 모든 프레임과 가장 유사한 프레임을 각 샷을 대표하는 키프레임으로 추출한다. 그 후, 비디오 씬 탐지를 위해, 컬러 히스토그램과 객체 컬러 히스토 그램에 해당하는 프레임의 시각적 특징을 기반으로 상향식 계층 클러스터링 방법을 이용하여 의미적인 연관성을 지니는 샷의 군집화를 통해 비디오 씬을 탐지하는 방법이다. 본 논문에서는 SDCEO의 프로토 타입을 구축하고 3개의 비디오 데이터를 이용한 실험을 통하여 SDCEO의 효율성을 평가하였고 샷 바운더리 식별의 성능의 정확도는 평균 93.3%, 비디오 씬 탐지 성능의 정확도는 평균 83.3%로 만족할만한 성능을 보였다.

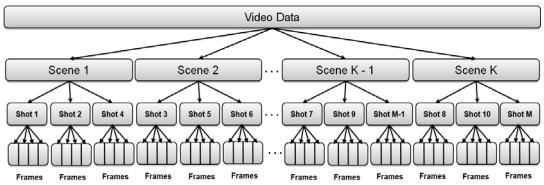
논문접수일 : 2012년 04월 06일 논문수정일 : 2012년 05월 14일 게재확정일 : 2012년 06월 04일
투고유형 : 국문일반 교신저자 : 최중민

1. 서론

비디오 데이터의 효율적인 관리 및 검색의 중요성은 강조되어 왔고(Gao et al., 2006; 이연호 외, 2011), 구조화되지 않은 복합 데이터의 형태를 지닌 비디오 데이터를 의미 있는 형태로 재구성하기 위한 비디오 파싱(video parsing)과 관련된 연구가 활발히 진행되었다(Lee, 2001).

비디오 데이터의 계층적인 구조는 <Figure 1>과 같다. 프레임(frame)은 비디오 데이터를 구성하는

가장 작은 단위로 픽셀(pixel)이 모여 이루어진 하나의 장면을 의미하고, 샷(shot)은 하나의 카메라에 기록된 연관성 있는 연속적인 프레임의 집합이고, 씬(scene)은 동일한 이벤트(event)를 구성하는 의미적인 연관성을 가지는 샷의 집합을 뜻한다.



<Figure 1> Hierarchical Structure of Video

초기 비디오 파싱에 대한 연구는 비디오 데이터를 샷 바운더리(shot boundary)로 분할하는데 초점을 맞추었다(Gargi, 2000). 하지만 물리적 바운더리인 샷 바운더리 탐지만으로는 의미적 연관성까지 고려하지 못하였다. 최근에는 동일한 이벤트에 속하는 의미적 연관성을 가진 샷을 군집화하여 의미적 바운더리인 씬(scene)으로 구조화하기 위한 연구가 활발히 진행되고 있다(Zhu and Liu, 2009; Mohonta, 2010). 기존의 연구는 그래프 모델을 이용하여 샷을 클러스터링(clustering)하거나(Yeung and Yeo, 1998; Rasheed and Shah, 2005), 샷 간의 유사도 평가를 통해 씬을 구성하고자 하였다(Hanjalic, 1999; Truong, 2003). 하지만 대부분의 기존 연구는 유사도 측정 시 색상의 유사도(Color similarity)만을 고려하였다. 색상의 유사도는 급진적 장면 전환(Hard Cut)과 같은 화면의 급격한 전환(Abrupt Transitions)을 탐지하는데 효율적이지만 디졸브(Dissolve), 페이드(Fade), 와이프(Wipe)와 같은 화면의 점진적인 전환(Gradual transitions)을 탐지하기는 어렵다(Huang et al., 2008).

이러한 문제점을 해결하기 위해, 본 논문에서는 프레임의 컬러 히스토그램(Color Histogram)과 코너 에지(Corner Edge), 그리고 객체 컬러 히스토그램(Object Color Histogram)에 해당하는 시각적 특징(Visual Feature)을 기반으로 비디오 씬을 탐지하는 방법(Scene Detector by using Color Histogram, Corner Edge and Object Color Histogram, SDCEO)을 제안한다. 우리는 비디오 데이터를 프레임 단위로 분할 후, 샷 바운더리 식별(Shot Boundary Identification)을 위해 컬러 히스토그램과 코너 에지의 유사도 비교를 통해 연관성 있는 일련의 프레임을 샷 바운더리로 병합하고, 각 샷 바운더리를 대표하는 키프레임(Key-Frame)을 추출한다. 그 후, 컬러 히스토그램과 객체 컬러 히스토그램 특징을 기반한 상향식 계층 클러스터링(Bottom-up Hierarchical Clustering)을 통해 동일한 이벤트를 구성하는 의미적인 연관성을 지니는 샷의 군집화를 통해 비디오 씬을 탐지한다. 또한 실제 비디오 데이터를 이용한 실험 및 분석을 통하여 SDCEO의 효율성을 평가하였다.

본 논문의 구성은 다음과 같다. 제 2장에서 SDCEO와 관련된 연구를 분석하고, 제 3장에서는 본 논문에서 제안하는 SDCEO의 시스템 구조에 대하여 기술한다. 제 4장에서는 시각적 특징을 기반한 샷 바운더리 식별 및 비디오 씬 탐지와 관련된 일련의 절차에 대하여 설명한다. 제 5장에서는 비디오 데이터를 이용한 실험을 통해 샷 바운더리 식별과 비디오 씬 탐지의 성능을 검증한다. 마지막으로 제 6장에서는 결론을 맺고 향후 과제를 제시한다.

2. 관련 연구

Chasanis et al.(2009)은 샷 단위로 분할된 비디오 데이터에서 fast global k-means를 기반한 spectral clustering 방법을 이용하여 키프레임을

추출하고, 키프레임간 색상의 유사도 평가를 통해 샷을 비디오 씬 단위로 군집화한다. 그 후, 비디오 씬 내 샷의 순차적 패턴(Sequential Pattern)을 분석하여 샷 레이블(Shot Label)을 생성하고, 이러한 패턴이 변화되는 경우 비디오 씬이 변경된다고 가정하여 비디오 씬을 분할하였다. 예를 들어, AB ABCBCABABC와 같은 샷으로 구성된 비디오 씬의 경우 A, B, C 샷이 비디오 씬의 샷 레이블이 되고, ABABCBDEDFEDEFDEDFEDEF와 같은 샷의 순차적 패턴은 ABABCB에 해당하는 첫 번째 씬과 DEDFEDEFDEDFEDEF에 해당하는 두 번째 씬으로 분할한다.

Yeung and Yeo(1996)은 비디오 콘텐츠의 색상 유사도와 시간적인 거리를 고려하여 샷을 클러스터링하는 time-constrained clustering 방법을 제안하였다. 제안하는 방법은 비디오 데이터를 샷으로 구성 후 키프레임을 선택한다. 그 후, 샷의 키프레임간 색상 특징이 유사하고 시간적인 거리가 임계치(threshold) 이하이면 하나의 씬으로 클러스터링한다.

Mohanta(Mohonta, 2010)은 명도 히스토그램 (Intensity Histogram), 웨이블릿 분해(Wavelet Decomposition), 에지 탐지(Edge Detection), 색상 상관도표(Color Correlogram)에 해당하는 시각적 특징의 유사도 평가를 기반으로 k-means 알고리즘을 수정한 클러스터링 방법을 이용하여 샷을 군집화한다. 그 후, Davies-Bouldin 인덱싱(Indexing) 방법을 기반한 클러스터 검증 분석 기법을 활용하여 최적의 클러스터 수를 만족할 때까지 반복적으로 클러스터링을 수행하여 비디오 씬을 탐지한다.

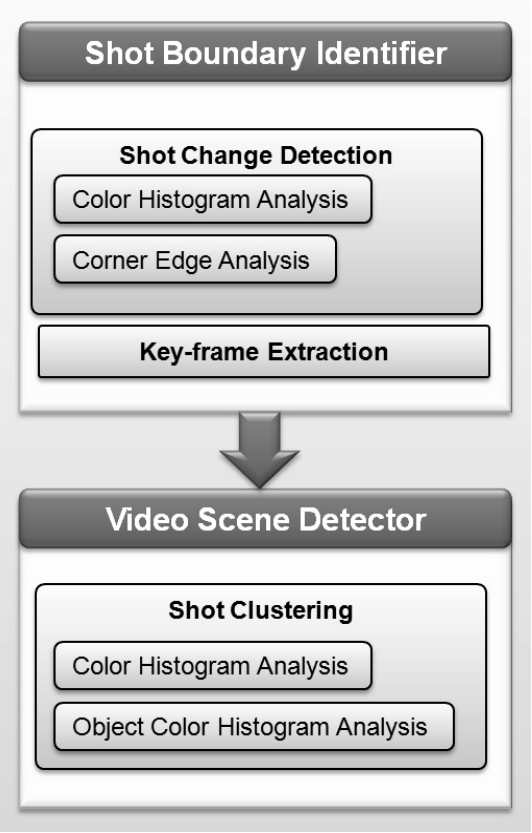
Zhu et al.(2009)는 비디오 데이터를 자동으로 비디오 씬으로 분할하는 새로운 방법을 제안하였다. Zhu가 제안한 방법은 rough-to-fine 알고리즘을 이용하여 샷을 탐지하고, 색상과 질감(Texture) 특징을 이용하여 각 샷에서 키프레임을 선택한 후, Template Matching 메소드(Method)를 통해 중복되는 키프레임을 제거한다. 최종적으로 샷 간의 시각적 유사도 평가와 시간적 거리 비교를 통해 연관된 샷을 하나의 씬으로 군집화한다.

위에서 언급한 기존의 연구들은 두 가지 문제점을 포함하고 있다. 첫째, 프레임간 혹은 샷간 유사도 측정 시 색상 정보만을 특징으로 고려하였다. 하지만 색상의 유사도는 점진적인 화면 전환 탐지가 어렵기 때문에 비디오 씬의 정확한 탐지를 보장할 수 없다. 둘째, 대부분의 기존 연구가 샷을 비디오 씬으로 군집화 시, 샷 간의 시간적인 거리를 함께 고려한다. 최근에는 교차 편집(다른 장소에서 동시에 일어나는 평행 행위를 시간상 전후 관계로 병치시키는 편집 기법) 등의 편집 기술이 널리 사용되고 있어 A→B→C→A→B→C와 같이 하나의 사건이 연속되지 않고, 여러 사건에 해당하는 샷이 번갈아 발생하는 경우가 빈번히 발생한다. 샷 간의 시간적인 거리를 고려한 경우, 교차 편집과 같은 비디오 편집 기법이 사용된 비디오 씬은 올바르게 탐지하지 못한다.

이러한 문제점을 해결하기 위해, 우리는 샷 간의 시간적인 거리 제약을 고려하지 않고 컬러 히스토그램을 이용하여 1차적으로 샷을 식별 후, 코너 에지 특징을 기반으로 카메라 관점의 변화 여부를 판단하여 샷을 정제한다. 그 후 컬러 히스토그램, 객체 컬러 히스토그램에 해당하는 시각적 특징을 기반으로 비디오 씬을 탐지하는 SDCEO을 제안한다.

3. 시스템 구조

SDCEO의 전체 시스템 구조는 <Figure 2>와 같다.



<Figure 2> The Structure of SDCEO System

*Shot Boundary Identifier*는 Shot Change Detection과 Key-frame *Extraction* 단계로 구성된다. ShotChangeDetection 단계는 프레임의 컬러 히스토그램 특징을 이용하여 1차적으로 프레임을 샷으로 구성하고, 1차적으로 구성된 샷을 코너 에지 특징을 기반으로 정제하여 최종적인 샷 바운더리를 식별한다. 그 후, *Key-frame Extraction* 단계에서 샷 바운더리를 대표하는 키프레임을 추출한다.

*Video Scene Detector*는 컬러 히스토그램과 객체 컬러 히스토그램 특징을 기반한 상향식 계층 클러스터링을 이용하여 연관된 샷 바운더리를 군집화함으로써 비디오 씬을 탐지한다.

4. 샷 바운더리 식별 및 씬 탐지

이 장에서는 비디오 데이터를 씬 단위로 분할하기 위한 샷 바운더리 식별 방법과 비디오 씬 탐지 기법에 대하여 설명한다.

4.1 샷 바운더리 식별

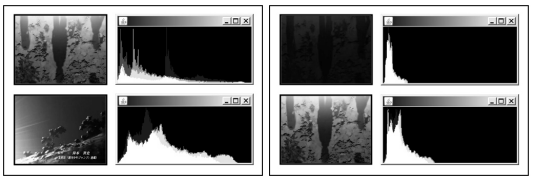
연관된 일련의 프레임을 하나의 샷으로 병합하기 위해 프레임의 컬러 히스토그램 분석을 통해 1차적으로 샷 바운더리를 구성한 후, 코너 에지 분석을 통해 최종 샷 바운더리를 식별한다.

4.1.1 컬러 히스토그램 분석

프레임의 컬러 히스토그램은 프레임 내 모든 픽셀의 양자화된 컬러(quantized color)의 비율을 표현한다(Lu et al., 2011). 컬러 히스토그램은 콘텐츠 기반 이미지 분석과 비디오 분석에서 좋은 성능을 보이기 때문에 널리 활용되는 특징이다(김광백 외, 2002; 허진경, 김향태, 2004). 샷 바운더리를 구성하기 위해, SDCEO는 각 프레임간 컬러 히스토그램 유사도 측정을 이용해 유사한 컬러 히스토그램을 가지는 연속적인 프레임을 샷으로 병합한다. 우리는 수식 (1)에 해당하는 히스토그램 유클리디안 거리(Histogram Euclidean Distance)(Sangoh, 2001)를 이용하여 순차적으로 연속된 프레임간 유사도를 측정한다.

$$sim(\alpha,\beta)=\sum_{r}\sum_{g}\sum_{b}(\alpha(r,g,b)-\beta(r,g,b))^2 \quad (1)$$

이때, α와 β는 각 프레임의 컬러 히스토그램을 의미하고, 컬러 히스토그램은 r(red), g(green), b(blue)의 색상 값을 포함한다.



(a) (b)

<Figure 3> An example of Color Histogram

컬러 히스토그램의 예는 <Figure 3>과 같다. <Figure 3(a)>는 하나의 샷에서 다른 샷으로 화면이 전환되는 예로 상이한 컬러 히스토그램을 보인다. 그러한 이유로 유사도 평가를 통해 서로 다른 샷으로 구별되고, <Figure 3(b)>는 동일한 샷에 속하는 프레임의 예로 유사한 컬러 히스토그램을 가지기 때문에 동일한 샷으로 결합된다.

4.1.2 코너 에지 분석

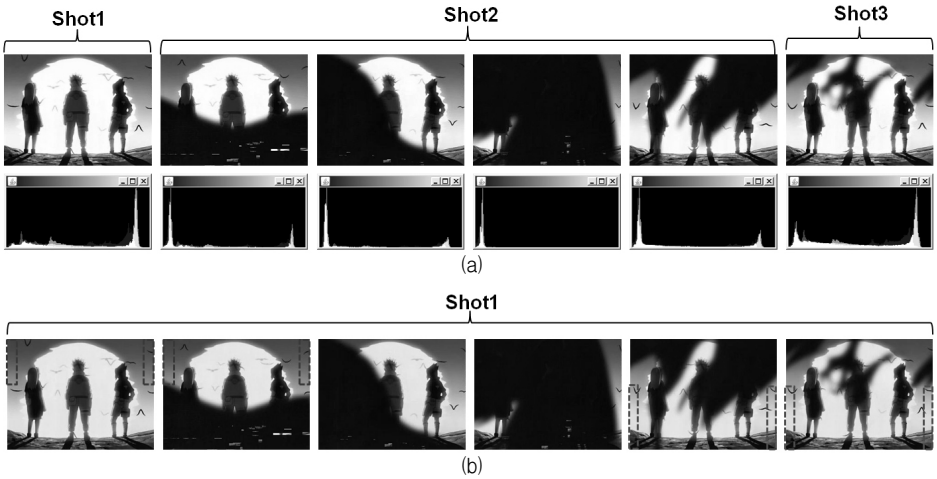
컬러 히스토그램은 카메라의 관점이나 이미지의 작은 변화 등에 민감하지 않은 이미지 비교에 적합한 특징으로 샷 바운더리 식별을 비롯한 많은 어플리케이션에서 널리 사용되는 방법이다(Manning and Raghavan, 2008). 하지만 위에서 언급했듯이 컬러 히스토그램 정보는 특정 사물의 개입 등으로 인해 급격하게 변화하기 때문에 색상만을 특징으로 하여 비디오 샷의 올바른 탐지는 어렵다. 이러한 이유로 실제 하나의 샷에 속해야 하는 프레임이 다수의 샷으로 분할되는 문제점이 발생한다. SDCEO는 다수의 샷으로 분할된 연관된 프레임을 하나의 샷으로 병합하기 위해 코너 에지 분석을 사용한다.

SDCEO의 코너 에지 분석의 목적은 컬러 히스토그램 특징을 기반으로 1차적으로 구성된 샷 바운더리를 정제하기 위한 것으로, 연속된 프레임의 배경이 되는 각 모서리의 상·하 쌍과 좌·우 쌍의 특정 구간이 동일한 경우 카메라의 관점이 변화하지 않고 유지된다고 가정할 수 있고, 카메라의 관점이 변화하지 않았다는 것은 '하나의 카메라에 기록된 연관성 있는 연속적인 프레임의 집합'이라는 샷의 정의에 따라 동일한 샷을 구성하는 프레임이라고 정의할 수 있다. 즉 이전 샷에 속하는 마지막 프레임과 다음 샷에 속하는 첫 프레임의 모서리의 특정 구간이 동일하다면 컬러 히스토그램에 포함된 노이즈나 특정 사물의 개입으로 인해 하나의 샷에 속해야 하는 프레임이 다수의 샷으로 분할되었다고 말할 수 있다. SDCEO는 이전 샷에 속하는 마지막 프레임과 다음 샷에 속하는 첫 프레임의 코너 에지의 상·하 쌍과 좌·우 쌍을 함께 비교하여 두 프레임의 코너 에지의 픽셀이 일치하는 특정 구간을 추출 후, 최대 길이를 가지는 구간을 두 프레임간의 유사도로 정의한다. 우리는 최대 구간이 임계치 이상일 경우 하나의 샷으로 간주하여 병합한다.

코너 에지를 이용한 샷 바운더리 정제 알고리즘은 알고리즘 1과 같다. *pF*와 *nF*는 코너 에지의 유사성을 비교할 대상프레임이고, getTopBottom

Algorithm 1 Refining Shot Boundary

```
1 : function RefineShotBoundary(pF, nF)
2 :     // pF, nF : compared frames with each other
3 :     pTopBottom[][] ← getTopBottomEdge(pF)
4 :     nTopBottom[][] ← getTopBottomEdge(nF)
5 :     pLeftRight[][] ← getLeftRightEdge(pF)
6 :     nLeftRight[][] ← getLeftRightEdge(nF)
7 :     tLeng ← getLength(pTopBottom, nTopBottom)
8 :     lLeng ← getLength(pLeftRight, nLeftRight)
9 :     maxLength ← compareLength(tLeng, lLeng)
10 :    return maxLength
11 : end function
```



<Figure 4> An Example of the Analysis of the Corner Edge

Edge(*frame*)과 getLeftRightEdge(*frame*)은 프레임의 상 · 하 쌍과 좌 · 우 쌍에 해당하는 코너 에지를 각각 가져오는 함수이다. getLength(*edge*1, *edge*2)함수는 두 코너 에지의 픽셀이 일치하는 최대 구간을 반환하고, compareLength(tLeng, lLeng) 함수는 두 구간 중 더 긴 구간을 반환한다.

코너 에지 분석의 예는 <Figure 4>와 같다. <Figure 4(a)>의 6개의 프레임은 동일한 샷 바운더리에 포함되어야 한다. 그러나 다른 객체의 출현에 의해서 컬러 히스토그램의 값이 급격하게 변화하였고, 그 결과로 3개의 샷 바운더리로 프레임이 분할되었다. 이러한 문제점은 이전 샷 바운더리와 다음 샷 바운더리 내 프레임의 코너 에지 특징 비교를 통해 <Figure 4(b)>와 같이 분할된 샷을 하나의 샷으로 병합 가능하다.

4.1.3 키프레임 추출

샷 바운더리를 식별한 후, SDCEO는 각 샷 바운더리에서 키프레임을 추출한다. 샷은 다수의 프레임으로 구성되어 있고, 샷 간의 비교를 위해 샷 내 모든 프레임을 비교하는 것은 비효율적이고 시간 소모적 작업이다. 이러한 이유로 기존의 많은 연구들이 샷을 대표하는 키프레임을 추출 후, 샷간의 비교를 키프레임간 비교로 대체하고자 하였다(Sakarya, Telatar, 2010; Amiri et al., 2011).

키프레임 선택을 위해, 우리는 수식 (1)의 히스토그램 유클리디안 거리를 이용하여 샷 바운더리 내 모든 프레임과 가장 유사한 프레임을 키프레임으로 선택한다.

4.2 비디오 씬 탐지

비디오 데이터는 다양한 이벤트로 구성되고, 각 이벤트는 연속적으로 발생하기도하지만 여러 사건이 번갈아가며 발생하기도 한다.

최근에는 평행편집과 같이 다른 두 시공간에서 벌어지는 상황을 번갈아가며 보여주는 편집 방식을 많이 사용한다. 이러한 이유로 비디오 데이터를 의미적 바운더리인 씬으로 구조화하는 것은 중요한 이슈이다.

우리는 컬러 히스토그램과 객체 컬러 히스토그램 특징을 기반한 상향식 계층 클러스터링을 통해 동일한 이벤트를 구성하는 의미적인 연관성을 지니는 샷의 군집화를 통해 비디오 씬을 탐지한다.

4.2.1 객체 컬러 히스토그램

객체 컬러 히스토그램은 프레임에서 객체(object)를 제외한 배경 색상(background color)을 모두 제거 후, 컬러 히스토그램을 추출한 것으로 유사한 프레임이 배경색에 의해 유사하지 않게 판단되거나 유사하지 않은 프레임이 배경색에 의해 유사하게 판단되는 경우를 감소시키기 위해 사용한다.

우리는 객체 컬러 히스토그램 추출을 위해 sobel edge detector(Sobel and Feldman, 1973)를 사용하여 프레임에서 객체의 에지을 추출하고, 프레임 이미지의 상×하×좌×우 각 모서리와 객체의 에지 사이의 배경색을 제거한다. 그 후 객체의 색상 정보만을 포함하는 프레임에서 컬러 히스토그램을 추출한다. 객체 컬러 히스토그램의 예는 <Figure 5>와 같다.

4.2.2 샷 클러스터링

우리는 의미적으로 연관된 샷을 Hierarchical Agglomerative Clustering(HAC) 방법을 이용하여 군집화한다. HAC 방법은 모든 클러스터가 하나로 병합될 때까지 클러스터의 쌍을 연속적으로 합쳐가는 방법이다(Manning and Raghavan, 2008). 우리는 각 샷을 Singleton Cluster로 간주 후, 키프레임의 컬러 히스토그램과 객체 컬러 히스토그램을 이용하여 클러스터링을 진행하면서 각 iteration마다 전체 샷 중 가장 유사한 두 샷을 병합한다. 이때 유사도 평가는 수식 (2)를 이용하여 행해진다.

$$shotSim(s_1, s_2) = \min_{x_1 \in s_1,\, x_2, \in s_2} distance(x_1, x_2)$$
$$distance(x_1, x_2) = \alpha \times cSim(x_1, x_2) + \beta \times oSim(x_1, x_2) \quad (2)$$

shotSim 함수는 각 샷을 구성하는 키프레임 중 가장 유사한 키프레임간 유사도를 계산하는 함수이다. *distance*는 키프레임간 유사도를 계산하는 함수이고, *cSim*과 *oSim*은 각각 수식 (1)을 이용하여 계산한 컬러 히스토그램과 객체 컬러 히스토그램의

<Figure 5> An Example of Extracting Object Color Histogram

유사도를 의미한다. 이때, α, β는 각각에 대한 가중치를 의미하고, 본 논문에서는 α, β에 동일한 가중치를 부여($\alpha = \beta = 0.5$)하였다.

SDCEO는 샷을 군집화한 후, 적절한 클러스터 수를 선택하기 위해 샷 바운더리간 유사 거리(similarity distance)가 임계치 h 이하일 때까지 반복적으로 클러스터링을 수행하여 최종 비디오 씬을 구성한다. 이때 하나의 씬으로 병합된 각 샷의 키프레임을 씬의 키프레임으로 선택한다.

5. 실험 평가

5.1 데이터 집합

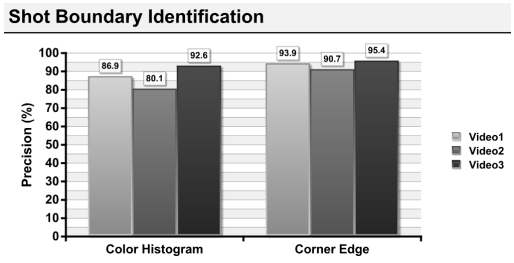
SDCEO의 효율성을 평가하기 위해, 우리는 다음과 같은 방법으로 실험을 위한 기준 데이터를 구성하였다. 주석이나 태그가 달리지 않은 약 6분 분량의 3개의 비디오 데이터에서 0.5초 단위로 프레임을 추출 후, 총 3명 중 2명이 수동으로 유사한 연속된 프레임을 샷으로 구성하고, 동일한 이벤트에 속하는 의미적으로 유사한 샷을 하나의 씬으로 구성하였다. 이때 의견 차이를 조절하기 위해 마지막으로 다른 한 명이 구축된 데이터를 검토하여 최종적으로 데이터를 구성하였다. SDCEO의 실험을 위해 구성된 데이터는 <Table 1>과 같다.

<Table 1> Date set for Performance Evaluation of SDCEO

File name	The number of frames	The number of shots	The number of scenes
Video1	714	107	19
Video2	815	161	21
Video3	799	100	20
Total	2328	368	60

5.2 성능 평가

본 논문에서는 SDCEO의 성능 평가를 위해, 샷 바운더리 식별과 비디오 씬 탐지 각각에 대한 성능을 평가하였다. 샷 바운더리 식별을 위한 컬러 히스토그램 분석의 임계치는 0.8로 설정하였고, 코너 에지 분석 시에는 프레임간 공유되는 픽셀이 존재하면 유사하다고 판단하였다. 또한 샷 클러스터링 시 임계치는 0.6으로 결정하였다. 샷 바운더리 식별 성능은 <Figure 6>과 같다.



<Figure 6> Experimental Results of Identifying Shot Boundary

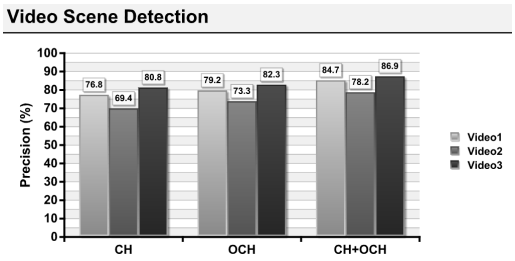
샷 바운더리 식별 시 컬러 히스토그램 분석만을 수행한 경우 평균 86.5%의 정확도를 보이고, 코너 에지 분석까지 모두 수행한 경우 평균 93.3%의 정확도를 보인다.

Video2의 컬러 히스토그램 분석 결과는 80.1%로 가장 낮은 성능을 보인다. 그 이유는 비디오 영상의 배경이 물 속, 물 수면, 하늘 등 유사한 컬러 히스토그램을 가지고, 영상의 좌우반전이나 다른 물체의 급격한 출현 등 노이즈가 되는 다양한 요소를 포함하고 있기 때문이다. 하지만 코너 에지 분석 후 90.7%의 성능을 보여 컬러 히스토그램 내 노이즈로 인해 올바르게 식별하지 못한 샷 바운더리를 어느 정도 보정하는 것을 확인할 수 있다.

Video3의 컬러 히스토그램 분석은 93.9%, 코너

에지 분석은 95.4%로 가장 높은 성능을 보인다. Video3은 각 사건별 등장하는 인물과 배경의 구분이 뚜렷하고, 다른 물체의 급격한 출현 등 노이즈가 되는 요소를 거의 포함하고 있지 않기 때문에 가장 높은 성능을 보였다. 컬러 히스토그램만을 이용한 경우에도 만족할만한 성능을 보이지만 코너 에지 분석 후, 다수의 샷으로 분할된 연관된 프레임을 하나의 샷으로 병합 및 정제하여 더 정확한 샷 바운더리를 식별하는 것을 확인할 수 있다.

비디오 씬 탐지 성능은 <Figure 7>과 같다. 성능 평가는 컬러 히스토그램 특징만 이용한 경우(CH), 객체 컬러 히스토그램 특징만 이용한 경우(OCH), 두 가지 특징을 모두 고려한 경우(CH+OCH)에 대하여 행하였다.



<Figure 7> Experimental Results of Detecting Video Scene

컬러 히스토그램만 이용한 경우 평균 75.7%, 객체 컬러 히스토그램만 이용한 경우 평균 78.3%, 두 가지 특징을 모두 고려한 경우는 평균 83.3%의 정확도를 보인다. 샷 바운더리 식별과 마찬가지로 Video2 데이터의 성능이 가장 낮고 Video3 데이터의 성능이 가장 높은 것을 확인할 수 있다. 이는 샷 바운더리 식별 결과가 비디오 씬 탐지 성능에도 영향을 미치기 때문이다. 컬러 히스토그램과 객체 컬러 히스토그램 특징을 모두 고려하여 비디오 씬을 탐지한 경우 만족할만한 성능을 보이는 것을 확인할 수 있다.

6. 결론

본 논문에서는 프레임의 컬러 히스토그램과 코너 에지, 객체 컬러 히스토그램에 해당하는 시각적 특징을 기반으로 샷 바운더리를 식별하고, 비디오 씬을 탐지하는 SDCEO를 제안하였다. 색상의 유사도 외에 에지 정보도 함께 활용하여 화면의 급격한 전환 뿐만 아니라 점진적인 전환까지 효율적으로 탐지한다는 점에서 주목할 만 하다.

SDCEO는 비디오 데이터를 프레임 단위로 분할 후, 컬러 히스토그램과 코너 에지 특징을 활용하여 샷 바운더리를 식별하고, 각 샷 바운더리를 대표하는 키프레임을 추출한다. 그 후, 컬러 히스토그램과 객체 컬러 히스토그램 특징을 기반한 상향식 계층 클러스터링을 통해 동일한 이벤트를 구성하는 의미적인 연관성을 지니는 샷의 군집화를 통해 비디오 씬을 탐지하였다. 또한 SDCEO의 성능 검증을 위해, 사람이 수동으로 구축한 데이터를 기준으로 하여 실험을 진행하였고 만족할만한 성능을 보였다.

본 논문에서 제안한 SDCEO의 문제점 중 하나는 코너 에지 분석과 객체 컬러 히스토그램 추출 시, 에지 정보를 추출하고 비교하는데 있어서 연산량과 연산시간이 많이 소요된다는 것이다. 따라서 향후 과제로 코너 에지 유사도 방식의 정제와 객체 컬러 히스토그램 추출 방식의 정제를 통해 SDCEO를 정제하고, 방대한 양의 데이터 집합을 기반한 실험을 통해 성능을 평가하고자 한다.

참고문헌

김광백, 윤홍원, 노영욱, “컬러 정보와 퍼지 C-means 알고리즘을 이용한 주차관리 시스템 개발”, *지능정보연구*, 8권 1호(2002), 87~101.

이연호, 오경진, 신위살, 조근식, “링크드 데이터를 이용한 협업적 비디오 어노테이션 및 브라우징 시스템”, *지능정보연구*, 17권 3호(2011), 203~219.

허진경, 김향태, “히스토그램 분포도 역추적 변경에 의한 영상 강조”, *지능정보연구*, 1권 8호(2004), 1~11.

Amiri, A., N. Abdollahi, M. Jafari, M. Fathy, “Hierarchical Key-Frame Based Video Shot Clustering Using Generalized Trace Kernel”, *Communicationsin Computer and Information Science*, Vol.241, No.5(2011), 251~257.

Chasanis, V., A. Likas, and N. Galatsanos, “Scene Detection in Videos Using Shot Clustering and Sequence Alignment”, *IEEE Transactionson Multimedia*, Vo1.11, No.1(2009), 89~100.

Gao, X., J. Li, and Y. Shi, “A Video Shot Boundary Detection Algorithm Based on Feature Tracking”, *In Proceedings of the Rough Sets and Knowledge Technology*, (2006), 651~658.

Gargi, U., R. Kasturi, and S. Strayer, “Performance characterization of video-shot-change detection methods”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.10(2000), 1~13.

Hanjalic, A., R. Lagendijk, and J. Biemond, “Automated high-level movie segmentation for advanced video-retrieval systems”, *IEEE Transactionson Circuitsand Systems for Video Technology*, Vol.9, No.4(1999), 580~588.

Huang, C., H. Lee, and C. Chen, “Shot Change Detection via Local Keypoint Matching”, *IEEE Transactionson Multimedia*, Vol.10, No.6(2008), 1097~1108.

Lee, M., Y. Yang, and S. Lee, “Automatic video parsing using shot boundary detection and camera operation analysis”, *Journal of the Pattern Recognition Society*, Vol.34, No.3(2001), 711~719.

Lu, H., Y. Tan, and X. Xue, “Real-Time, Adaptive, and Locality-Based Graph Partitioning Method for Video Scene Clustering”, *IEEE Transactionson Circuitsand Systems for Video Technology*, Vol.21, No.11(2011), 1747~1759.

Manning, C. and P. Raghavan, H. Schutze, “Introduction to Information Retrieval”, Cambridge University Press, 2008.

Mohonta, P., S. Saha, and B. Chanda, “A Hueristic Algorithm for Video Scene Detection Using Shot Cluster Sequence Analysis”, *In Proceedingsof the 7th Indian Conferenceon Computer Vision, Graphicsand Image Processing*, (2010), 464~471.

Pass, G., R. Zabih, and J. Miller, “Comparing Images Using Color Coherence Vectors”, ACMConferenceon Multimedia, (1996), 65~74.

Rasheed Z. and M. Shah, “Detection and representation of scene in videos”, *IEEE Transactionsof Multimedia*, Vol.7, No.6(2005), 1097~1105.

Sakarya, U., Z. Telatar, “Video scene detection using graph-based representations”, *Signal Processing Image Communication*, Vol.25, No.10(2010), 774~783.

Sangoh, J., “Histogram-Based Color Image Retrieval”, Technical Report, Psych221/EE362 Project, 2001.

Sobel, I. and G. Feldman, “A 3x3 Isotropic Gradient Operator for Image Processing”, *InProceedings Pattern Classification and Scene Analysis*, (1973), 271~272.

Truong, B., S. Venkatesh, and C. Dorai, “Scene extraction in motion picture”, *IEEE Transactionson Circuitsand Systems for Video Technology*, Vol.13, No.1(2003), 5~15.

Yeung, M. and B. Yeo, “Segmentation of video

by clustering and graph analysis", *Journal of Computer Visionand Image Understanding*, Vol.71, No.1(1998), 97~109.

Yeung, M. and B. Yeo, "Time-Constrained Clustering for Segmentation of Video into Story Units", *InProceedings of 13th International Conference on Pattern Recognition*, Vol.3(1996), 375~380.

Zhu, S. and Y. Liu, "Video scene segmentation and semantic representation using a novel scheme", *Multimedia Tools and Applications*, Vol.42, No.2(2009), 183~205.

Abstract

Video Scene Detection using Shot Clustering based on Visual Features

Dongwook Shin* · Taehwan Kim** · Joongmin Choi***

Video data comes in the form of the unstructured and the complex structure. As the importance of efficient management and retrieval for video data increases, studies on the video parsing based on the visual features contained in the video contents are researched to reconstruct video data as the meaningful structure. The early studies on video parsing are focused on splitting video data into shots, but detecting the shot boundary defined with the physical boundary does not cosider the semantic association of video data. Recently, studies on structuralizing video shots having the semantic association to the video scene defined with the semantic boundary by utilizing clustering methods are actively progressed.

Previous studies on detecting the video scene try to detect video scenes by utilizing clustering algorithms based on the similarity measure between video shots mainly depended on color features. However, the correct identification of a video shot or scene and the detection of the gradual transitions such as dissolve, fade and wipe are difficult because color features of video data contain a noise and are abruptly changed due to the intervention of an unexpected object.

In this paper, to solve these problems, we propose the Scene Detector by using Color histogram, corner Edge and Object color histogram (SDCEO) that clusters similar shots organizing same event based on visual features including the color histogram, the corner edge and the object color histogram to detect video scenes. The SDCEO is worthy of notice in a sense that it uses the edge feature with the color feature, and as a result, it effectively detects the gradual transitions as well as the abrupt transitions.

The SDCEO consists of the Shot Bound Identifier and the Video Scene Detector. The Shot Bound Identifier is comprised of the Color Histogram Analysis step and the Corner Edge Analysis

* Department of Computer Science and Engineering, Hanyang University
** BK21 AIS Team, Hanyang University
*** Corresponding Author: Joongmin Choi
Department of Computer Science and Engineering, Hanyang University
1271, Sa3-Dong, Sangnok-Gu, Ansan, Gyeonggi-Do 426-791, Korea
Tel: +82-31-400-4110, Fax: +82-31-409-7351, E-mail: jmchoi@hanyang.ac.kr

step. In the Color Histogram Analysis step, SDCEO uses the color histogram feature to organizing shot boundaries. The color histogram, recording the percentage of each quantized color among all pixels in a frame, are chosen for their good performance, as also reported in other work of content-based image and video analysis. To organize shot boundaries, SDCEO joins associated sequential frames into shot boundaries by measuring the similarity of the color histogram between frames. In the Corner Edge Analysis step, SDCEO identifies the final shot boundaries by using the corner edge feature. SDCEO detect associated shot boundaries comparing the corner edge feature between the last frame of previous shot boundary and the first frame of next shot boundary. In the Key-frame Extraction step, SDCEO compares each frame with all frames and measures the similarity by using histogram euclidean distance, and then select the frame the most similar with all frames contained in same shot boundary as the key-frame.

Video Scene Detector clusters associated shots organizing same event by utilizing the hierarchical agglomerative clustering method based on the visual features including the color histogram and the object color histogram. After detecting video scenes, SDCEO organizes final video scene by repetitive clustering until the simiarity distance between shot boundaries less than the threshold h.

In this paper, we construct the prototype of SDCEO and experiments are carried out with the baseline data that are manually constructed, and the experimental results that the precision of shot boundary detection is 93.3% and the precision of video scene detection is 83.3% are satisfactory.

Key Words : Video Scene Detection, Shot Clustering, Shot Boundary Identification, Video Parsing, Visual Feature Extraction

저 자 소 개

신동욱

경원대학교 인터넷미디어학과를 졸업하였고, 한양대학교 대학원 컴퓨터공학과에서 석사학위를 취득하였다. 현재 한양대학교 대학원 컴퓨터공학과 박사과정 중이다. 관심 분야는 소셜 네트워크, 데이터 마이닝, 정보검색/정보추출, 웹지능, 인공지능, 지능형 모바일정보시스템 등이다.

김태환

인천대학교 컴퓨터공학과를 졸업하였고, 한양대학교 대학원 컴퓨터공학과에서 석사학위를, 한양대학교 대학원 컴퓨터공학과에서 공학 박사학위를 각각 취득하였다. 현재 한양대학교 BK21 AIS 사업단에서 Post-Doc로 재직 중이다. 한국 정보과학회, 정보처리학회, 지능정보시스템학회, 인터넷정보학회 등의 정회원(혹은 준회원)이며, 관심분야는 웹지능, 텍스트마이닝, 정보검색/정보추출, 인공지능, 상황인지 등이다.

최중민

서울대학교 컴퓨터공학과를 졸업하였고, 서울대학교 대학원 컴퓨터공학과에서 석사학위를, 미국 State University of New York at Buffalo에서 컴퓨터학 박사학위를 각각 취득하였다. 1993년부터 1995년까지 한국전자통신연구원(ETRI)에서 선임연구원으로 재직하였으며, 1995년부터 현재까지 한양대학교 컴퓨터공학과 교수로 재직 중이다. 한국 정보과학회, 정보처리학회, 지능정보시스템학회, 인터넷정보학회, 미국 IEEE, ACM 등의 정회원이며, 관심분야는 웹지능, 텍스트마이닝, 정보검색/정보추출, 인공지능, 지능형 모바일정보시스템 등이다.