

시맨틱 검색을 위한 디지털 비디오 라이브러리 구축에 관한 연구

A Study on Digital Video Library Development for Semantic- Sensitive Retrieval

장 상 현* · 임 석 종**

Sang-Hyun Jang · Seok-Jong Lim

차 례

- | | |
|-----------------------|---------------------|
| 1. 서 론 | 4. 시맨틱 검색을 위한 장면 분류 |
| 2. 자동색인을 위한 비디오 분할 방법 | 5. 결 론 |
| 3. 분할된 비디오 분석 방법 | • 참고문헌 |

초 록

디지털 비디오 압축과 인터넷 기술의 발전으로 비디오에 대한 수요가 폭발적으로 늘어나고 있으며, 사용자 제작 콘텐츠(UCC)도 대량으로 생산되고 있다. 따라서, 디지털 비디오 라이브러리 구축과 검색에 대한 기술 연구가 어느 때보다 시급하게 요구되고 있다. 그러나 사용자가 전체 비디오의 일부 내용을 의미기반으로 검색해서 보고 싶은 욕망을 충족시켜주는 것은 매우 어려운 상황이다. 본 고에서는 디지털 비디오를 의미 검색하기 위하여 장면을 뽑아내고, 내용을 분석하는 방법을 제안하고, 5개의 스포츠(축구, 야구, 골프, 농구, 배구) 뉴스 비디오를 자동 분류한 실험 결과를 소개하였다.

키 워 드

디지털 도서관, 시맨틱, 비디오, 자동 색인

* 한국교육학술정보원 교육정보화센터 책임연구원
(Principal Researcher, Education Information Center, Korea Education and Research Information Service, shjang@keris.or.kr)

** 한국과학기술정보연구원 지식기반팀 선임연구원
(Senior Researcher, Knowledge Asset Team, Korea Institute of Science and Technology Information, seoklim@kisti.re.kr)

• 논문접수일자 : 2006년 11월 16일

• 게재확정일자 : 2006년 12월 13일

ABSTRACT

With the advancement of internet and video compression technology, there has been an increasing demand for video, and produced a large quantity contents of UCC. Therefore, Semantic-sensitive retrieval and construction for digital video library is more in demand than ever. However, it is extremely difficult to categorize and label scenes in any video automatically for searching wanted scene. This study proposes a method to extract certain scenes and analyze the video content, and shows the experimental results after categorizing 5 sports news(soccer, baseball, golf, basketball, and volleyball).

KEYWORDS

Digital Video Library, Semantic Retrieval, Automatic Indexing, Annotation

1. 서 론

디지털 동영상은 압축과 인터넷 기술의 발전으로 그 수요가 폭발적으로 늘어나고 있으며, 이용자는 동영상의 내용 정보를 직접 검색하려는 요구가 증가하고 있다. 동영상 이용자들은 영화나 방송의 특정 장면을 보고자 하는 일반적인 욕구가 있다. 이러한 욕구를 해결하기 위한 연구가 수십 년간 진행되었고, 이러한 연구를 위한 주요 기술로써 비디오 색인에 대한 연구가 이루어지고 있다. 디지털 비디오 라이브러리 구축의 관점은 비디오 색인이다. 디지털 비디오의 색인 정보는 문서식별자, 내용 기술 그리고 키워드로 구성된다. 수작업에 의한 키워드 할당은 디지털 비디오 라이브러리 데이터베이스의 시간적이고 공간적인 특성으

로 인해 실질적인 방법은 아니다. 또한, 개별 목록자의 시각이 시간에 따라 다양하며, 이변트적인 이용자의 요구와는 차이가 있다. 디지털 비디오 라이브러리 구축에서 일반적인 내용 기반 검색 기술은 이미지 인식기법을 포함한 분할 방법(segmentation method), 비디오 분할을 위해 광범위한 폐쇄된 캡션 파일(또는 방송 대본)의 분석, 폐쇄된 캡션 파일이 없는 경우에 단어 지시자를 사용한 비디오 클립 색인이 있다(Wei Li et al, 1996). 본 논문에서는 동영상의 이미지 인식 기법을 활용한 디지털 동영상 검색에 국한하여 살펴보고자 한다.

동영상의 내용을 기반으로 디지털 비디오를 검색하기 위해서는 비디오의 장면을 분류하고 장면에 의미를 부여하는 것이 필요하다. 의미 부여를 자동화하기 위해서는 비디오에 포함된

소리와 영상의 이해 및 분석을 통해 가능하다 (Janko Calic and Ebroul Izquierdo 2002; Yueting Zhuang, Ruogui Xiao, and Fei Wu 2003).

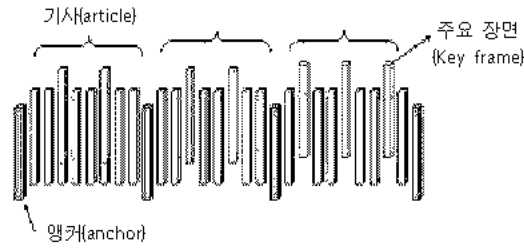
최근 인터넷 포털에서도 이미지와 동영상에 대한 검색을 강화하고 있으나, 그 방법이 파일 단위의 속성을 수동으로 입력한 메타데이터 (metadata)나 방송 대본(script)과 같이 문자(text)를 기반으로 검색한다. 그러나 동영상이 지닌 특징인 색깔(color), 움직임(motion), 무늬(texture), 소리(audio) 등을 활용하여 내용을 자동으로 분류함으로써 동영상의 내용을 검색할 수 있다.

보편적으로 이용되는 동영상은 영화, 드라마, 스포츠, 다큐멘터리, 뉴스, 교양, 오락 등 다양한 장르가 있다. 이 중 대부분이 비정형화된 형태의 동영상으로써 자동 색인이 매우 어렵지만 뉴스와 같이 정형화된 동영상은 자동 장면 분류가 가능하다(SangHyun Jang and HyungJe Cho 1997). 이런 이유로 뉴스 비디오는 자동화된 비디오 색인을 통해 원하는 소식을 찾는 주문형 뉴스(News On Demand) 시스템으로 발전하고 있다(H. J Zhang, C. Y. Low, S. W. Smoliar, and J. H. Wu 1995; SangHyun Jang and HyungJe Cho 1997).

이와 관련된 연구를 살펴보면, Gong's work는 TV 축구경기를 분석하기 위해서 텍스처와 객체의 색깔을 이용했다. 또 다른 연구에서는 경기장면을 분류하기 위해 카메라 움직임과 선수의 위치를 이용하여 FSM(Finite State

Machine)을 이용하여 축구비디오에서 하이라이트 장면을 뽑아냈다. Hanjalic은 소리, 움직임, 그리고 샷 특징(shot features)를 이용해서 축구의 하이라이트를 찾아냈다. Xie 등은 먼저 HSV(Hue Saturation Value) 컬러 모델을 이용해서 내부 비율(dominant ratio)를 찾아내고 매크로 블럭의 움직임 정도를 계산한다. 마지막으로 운동장면과 휴식장면을 HMM(Hidden Markov Model)을 이용하여 분리한다. Duan and Tian 등은 하위 특징, 중간 표현, 상위 이벤트를 포함하는 정보를 가지고 있고, 스포츠 비디오의 이벤트 찾기, 하이라이트 추출, 요약과 개별화를 위해 사용하는 중간 수준의 프레임워크(mid level framework)을 제안했다.(Shuqiang Jiang, Qixiang Ye, Wen Gao, and Tiejun Hunang 2004)

본 연구는 하나의 스포츠 비디오를 분석하는 수준에서 나아가 뉴스 비디오에서 스포츠를 분류(classification)하기 위한 방법을 제안한다. 이를 위해 뉴스 비디오의 특징인 앵커를 장면 전환의 첫 프레임으로 지정하여 장면을 구분하고 장면내의 대표 프레임의 경기장 배경 색깔을 특징으로 원 거리 장면(long view) 화면을 찾아 1차 스포츠 종류를 구분하고 화면의 선수와 공의 움직임 정보를 활용하여 2차 스포츠 분류를 한다. 실험을 위하여 축구, 야구, 골프, 농구, 배구 5종의 스포츠를 분류한다.



〈그림 1〉 뉴스비디오의 구조

2. 자동 색인을 위한 비디오 분할 방법

CNN 이나 KBS와 같은 뉴스 비디오는 하루의 사건 사고를 담고 있으므로 정치, 경제, 사회 등 매우 다양한 기사를 담고 있다. 그 기사 내용을 자동으로 분석하여 의미적으로 분석하는 것은 쉬운 일이 아니다. 일반적으로 비디오의 내용을 분석하는 방법은 1) 샷의 특징을 이용하는 샷 기반(shot based) 방법과 2) 비디오의 객체(object or region) 기반으로 객체의

속성을 이용하는 방법으로 분류된다(M.R. Naphade and T.S. Huang 2001, 3: 141-151). 그러나 일반 뉴스 비디오는 정형화된 구조를 가지고 있지 않으며 그 내용도 다양하여 자동화된 분석은 거의 불가능하다. 뉴스 비디오는 〈그림 1〉과 같이 앵커화면과 기사화면으로 구성된 특징과 스포츠 경기의 경기장 특징을 혼합하여 사용하면 스포츠 뉴스를 의미적으로 분류하는 것이 가능하다(H. J Zhang, C. Y. Low, S. W. Smoliar, and J. H. Wu 1995; SangHyun Jang and HyungJe Cho

기사 번호	아니운서 화면 시작/종료 프레임	스포츠 장면 추출된 키 프레임
No.1		
No.2		
No.3		
No.4		

〈그림 2〉 장면 분할 결과

1997).

먼저, 장면 추출은 앵커를 찾고 기사 대표 프레임(key frame)을 찾는 것으로 앵커화면을 경계로 활용하고 대표 프레임은 샷으로써 스포츠 내용을 구분하는 데 사용한다. 일반적으로 장면 추출은 동영상의 시간적(temporal), 공간적(spatial) 특징을 활용하는데, 먼저 시간적 특징을 이용하여 이웃 프레임과 색상 비교를 하여 그 차가 임계값(threshold) 이상이면 대표 프레임으로 선정한다. 선정된 대표 프레임 간의 차이(distance)는 앵커 화면을 찾는 방법으로 이웃과 변화도가 큰 경우이므로 임계값 이하일 때 앵커화면으로 추측할 수 있다 (SangHyun Jang and HyungJe Cho 1997). <그림 2>는 스포츠 뉴스의 앵커 화면과 스포츠 장면의 대표 프레임을 추출한 결과이다.

3. 분할된 비디오 분석 방법

3.1. 경기장의 색상 분석

스포츠 경기장은 스포츠마다 형태나 색상이 서로 다른 특성을 가지고 있다. <그림 3>의 스포츠별 경기장을 살펴보면, (a)와 (b)는 각각 축구와 골프의 필드로 주로 초록색이고, 야구 필드는 초록색, 흙색(주황색)이거나 (c)와 같이 두 가지 색상을 포함하기도 한다. (d)는 배구로 연분홍색의 필드를 가지고, (e)의 농구 필드는 노란색(주황색) 등의 특징을 가진다. 따라서 스포츠를 분류하는 데 있어서 필드의 색상은 중요하고 유용한 특징 정보라고 할 수 있다(Lei Wang, Boyi Zeng, Steve Lin, Guangyou Xu, and Heung Yeung Shum 2004).

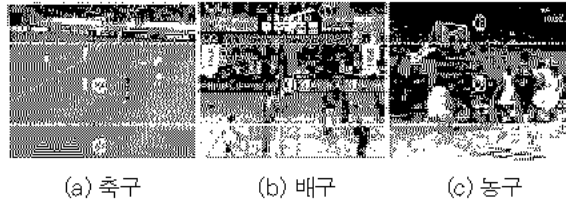
스포츠 경기장의 색상을 분석하기 위해서,



<그림 3> 스포츠 경기 장면 예



<그림 4> 스포츠 경기 장면 예



〈그림 5〉 우세색 검출을 위한 영역 분할

우선 스포츠 장면에서 추출된 키프레임들의 영상처리(Image Processing) 과정이 이루어진다. 영상처리의 목적은 영상에서의 불균일한 색상을 부드럽게 만들어 색상의 분포도를 균일화하고, 또한 색상수의 분포도를 줄여서 주변의 유사 색상들을 동일한 색상으로 그룹화하기 위한 것이다.

키프레임의 영상처리 과정을 거친 후, 스포츠별 필드를 식별하기 위해 〈그림 4〉와 같은 영상을 3부분의 영역으로 분할하여 영역별 색상의 분포도를 분석한다.

영상의 색상 분포도의 분석은 3영역별로 세분화하여 우세색(Dominant Color)을 검출한다. 3영역의 분할 목적은 〈그림 5〉의 예와 같이 스포츠의 영상을 살펴보면, 1영역은 대부분 필드와 상관없는 관중이나 광고판 등의 스포츠 필드 색상의 식별에 불필요한 요소들을 많이 포함하고 있으며, 3영역은 자막으로 인해

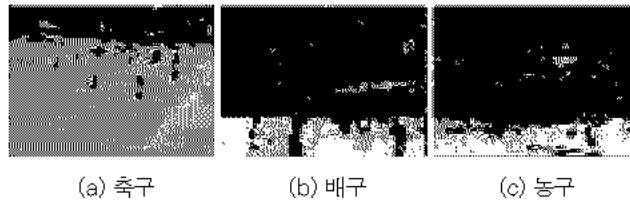
필드를 가리므로 필드 색상을 파악하기 어렵게 한다.

또한 〈그림 5〉의 예에서 축구나 골프인 경우는 (a)와 같이 1, 2, 3영역 모두 동일한 색상을 띠고 있어서 우세색을 식별하기 쉽지만, (b)와 (c)의 경우는 전체영상에 대해서 동일 색상의 분포가 아니므로 우세색을 검출하기 어렵다.

따라서 필드 색상을 분석하기 위해서, 추출된 키프레임의 영상에 대해 〈그림 5〉처럼 1, 2, 3영역별 우세색을 검출하여 종합적으로 분석하는데 이를 원거리시야(long view)와 부분시야(partial view)라 칭한다. 원거리시야(long view) 분석이란 카메라가 경기장 필드를 최대한 포함할 수 있도록 먼 거리에서 영상을 포착하는 경우로 1, 2, 3영역에 대해서 모두 동일한 우세색을 나타내는 것을 말한다. 부분시야(partial view) 분석은 농구나 배구와 같이 필드의 색상이 전체적으로 나타나지 않는 경우에

〈표 1〉 경기장 색깔과 스포츠의 관계

우세색(Dominant Color)	Hue	스포츠
초록색	65 <- H <- 110	축구, 골프, 야구
노랑색(주황색/흙색)	40 <- H <- 55	농구, 야구
연분홍색	20 <- H <- 35	배구



〈그림 6〉우세색 검출 결과

대해서 필드 색상을 분석할 수 있도록 키프레임의 일부 영역인 2영역이나 3영역의 영역별 우세색을 검출하는 것이다.

스포츠별로 필드 색상을 구별하기 위해서 키프레임의 영상처리 과정 후에 HSV상의 색상으로 변환하여 〈표 1〉과 같이 정의된 Hue값의 범위에 따라 스포츠별 필드의 우세색을 검출한다. 〈표 1〉에 정의된 Hue값의 범위에 따라 필드의 우세색 검출 결과가 〈그림 6〉이다.

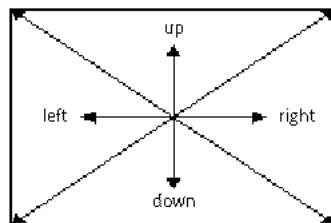
3.2. 모션 분석(Analysis of Motion)

스포츠 비디오 영상에서 카메라나 객체의 움직임 방향은 스포츠를 구별하는데 유용한 특징정보이다. 즉, 스포츠별 모션 방향성을 살펴보면, 필드 색상이 비슷한 축구와 골프에서 축

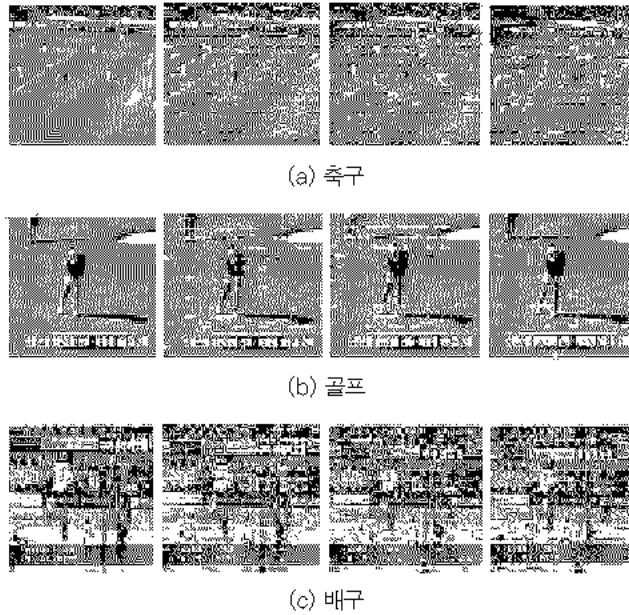
구는 한쪽 방향인 왼쪽이나 오른쪽 방향으로 움직임을 보이고, 골프는 거의 움직임 없다. 또한 배구의 경우는 네트를 중심으로 왼쪽과 오른쪽 방향의 움직임을 동시에 보이거나 공의 방향에 따라 왼쪽 또는 오른쪽 방향의 움직임을 보이고, 농구의 경우는 공의 방향에 따라 위쪽 방향성을 보이거나 불규칙하게 나타나고, 야구는 불규칙한 방향성을 가진다.

스포츠 영상에서 모션 방향성의 분석은 MPEG4 인코딩 처리 과정에서 모션 벡터 값을 추출하여 매크로 블록의 x축과 y축의 모션 벡터의 평균값을 모션의 방향성의 크기로 산출한다. 따라서 모션 방향의 크기는 X, Y방향에 대한 모션 벡터(motion vectors)의 총 수이다.

그 결과 모션 방향성의 크기를 통해서 〈그림 7〉과 같이 4방향에 대한 모션의 방향을 결



〈그림 7〉화면의 모션 정보



〈그림 8〉모션 정보 검출 결과

정한다.

추출된 모션의 방향을 표현한 결과가 〈그림 8〉이고, 모션 움직임의 분석은 추출된 스포츠 샷에 대해 연속 프레임들이 동일한 방향성을 가지고 있는지를 검출하여 스포츠 장르를 분류한다.

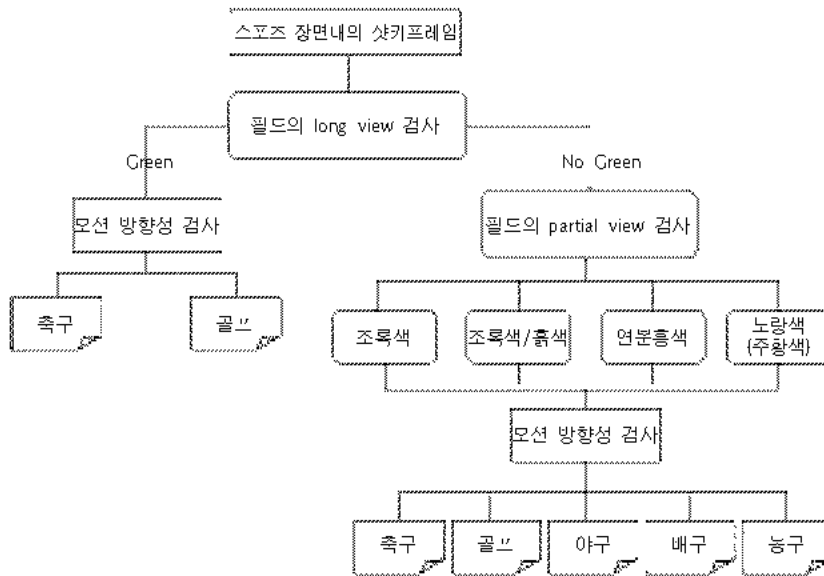
축구(a)는 공의 흐름에 따라 여러 프레임이 동일하게 오른쪽 방향을 보이고, 골프(b)는 거의 움직임이 없고, 배구(c)는 네트를 중심으로 양쪽 방향이 동시에 보인다.

4. 시맨틱 검색을 위한 장면 분류

의미기반 비디오 검색 방법은 (a) 룰 기반(Rule based approach) 접근과 (b) 통계적 접근

(Statistical approach)으로 분리할 수 있다. 뉴스와 같이 의미 단위(semantic unit)가 잘 정의된 구조는 룰 기반 접근을 많이 사용한다(Sohaib Khan and Mubarak Shah 2001; Zhibin Lei, Wu Chou, Jialin Zhong, and Chin Hui Lee 2000). 제안한 분류 방식은 스포츠 뉴스 비디오에서 추출된 스포츠 장면의 장르는 우세색, 모션 방향성의 특징 정보를 가지고 〈그림 9〉 분류 과정을 통해 결정한다.

스포츠 장면내의 키프레임들에 대해서 우세색을 검출한다. 검출된 우세색을 기반으로 원거리시야(long view)를 추출하여 우세색이 초록색인 경우와 초록색이 아닌 경우로 분류한다. 원거리시야(long view)에서 우세색이 초록색인 경우에 대해서 모션 방향성을 검사하여



〈그림 9〉 스포츠 장면 분류 과정

축구와 골프로 분류한다. 원거리시야(long view)의 우세색이 초록색이 아닌 경우는 부분시야(partial view) 검사를 통해서 각 스포츠별 필드의 필드 색상을 검출하고, 이에 대한 모션 방향성을 검출하여 스포츠 장르를 분류한다.

실험은 Pentium4 급 PC에서 Windows의 운영체제, Microsoft Visual C++ 6.0의 환경에서 이루어졌다.

입력 영상은 축구, 배구, 야구, 농구, 골프 등 5종의 스포츠 뉴스 비디오 영상의 AVI형식으로 영상의 해상도는 320 x 240를 표준 크기로 하고, 기타 해상도 크기에 대해 제약을 받지 않는다.

입력 영상의 제약 조건으로는 반드시 맨 처음과 스포츠 장면 사이에는 앵커장면이 반드시 포함되어야 있어야 한다. 처음의 앵커 장면을

기반으로 스포츠 장면간의 앵커 장면을 추출하므로 앵커는 동일한 위치에 있어야 한다. 또한 스포츠 장면이 아닌 앵커와의 인터뷰 등의 장면은 삭제하여야 앵커 장면과 스포츠 장면을 구별하여 검출할 수 있다.

본 실험은 스포츠 장르를 구별하기 위해서 스포츠 비디오의 총 25 개, 87개 장면수의 기본 학습 데이터(training mode)를 통해 스포츠별 기준치를 찾고 이를 바탕으로 스포츠 장르 분류의 정확도를 알아보기 위해서 테스트 데이터(test mode)로 스포츠 비디오 총 45개, 154개 장면에 대해 스포츠 장르의 분류 정확도를 측정하였다.

〈표 2〉에서 스포츠 장르별 분석 결과를 살펴보면, 축구와 야구는 정확하게 검출되는 것을 볼 수 있고, 배구와 농구도 필드 색상과 모션의

〈표 2〉 (a) 실험 자료

	축구	골프	야구	농구	배구	기타	장면수
축구	18	6			1		25
골프	6	3	3				12
야구	3		24				27
농구				15		1	16
배구				1	6		7
정확도	18	3	24	15	6		66/87(76%)

〈표 2〉 (b) 분류 결과

	축구	골프	야구	농구	배구	기타	장면수
축구	50	4	2				56
골프	12	4	3				19
야구	3	4	25			1	33
농구			1	17	5	2	25
배구				3	18		21
정확도	50	4	25	17	18		114/154(74%)

방향성에 따라 정확하게 검출 결과값을 얻었다. 골프인 경우는 우세색 검출을 바탕으로 방향성 검사에서 방향의 변화가 많이 나타나므로 정확도가 떨어짐을 알 수 있다.

이 연구에서 스포츠 뉴스 비디오의 전체 정확도는 학습 데이터와 테스트 데이터에 대한 전체 장면수는 241개이고 정확하게 분류한 개수는 180개로 75%의 정확도를 얻었다.

5. 결 론

본 연구에서는 스포츠 뉴스 비디오의 장면

을 분류하기 위한 하나의 시도로서 영상의 특징을 색상과 움직임을 이용하여 70~75%의 분류 정확도를 얻었다. 그러나 보다 정확도를 높이기 위해서는 영상이 가지고 있는 추가적인 특징 요소를 사용해야 하는 데 이들 요소로는 일반적으로 무늬(texture), 카메라 이동(camera motion), 객체(object)의 모양(shape) 등이 이용될 수 있다.

스포츠 뉴스 비디오에 대한 추가연구를 수행할 경우, 추가로 고려될 수 있는 특징으로는 근접시야(close view)에 대해 선수 객체를 추출하여 스포츠별로 서로 다른 유니폼 특징을

반영하여 분류한다면 정확도를 높일 수 있을 것이다.

또한, 정확한 분류를 위해 오류에 대한 피드백을 활용하여 특징의 가중치를 조정하는 방법과 객체를 추출하여 특징으로 활용하는 분야에 대한 연구가 추가된다면 그 정확도와 활용도가 더 높아질 것으로 예상된다.

참고문헌

- H. J Zhang, C. Y. Low, S. W. Smoliar, and J. H. Wu. 1995. "Video Parsing, Retrieval and Browsing: An Integrated and Content Based Solution". New York: ACM.
- Jame Ze Wang, Jia Li, Desmond Chan, and Gio Wiederhold. 1999. "Semantics sensitive Retrieval for Digital Picture Libraries", *D Lib Magazine*.
- Janko Calic and Ebroul Izquierdo. 2002. "Efficient Key Frame Extraction and Video Analysis". Los Alamitos: IEEE.
- Ichiro IDE, Reiko HAMADA, and Shuichi SaKAI. 2001. "An attribute based news video indexing". New York: ACM.
- J. Assfalg, M. Bertini, C. Colombo, and A. Del Bimbo. 2001. "Feature Extraction and Content Analysis for Sports Videos Annotation", *Multimedia Information Retrieval Ottawa Canada, New York: ACM*.
- Jiaping Fan, Ahmed K. Elmagarmid, Xingquan Zhu, Walid G.Aref, and Lide Wu. 2004. "Hierarchical Video Shot Classification, Indexing, and Accessing". IEEE, *Transaction on Multimedia*, 6(1).
- Jiaping Yang. 2003. "Content Based Video Identification: A Survey". Los Alamitos: IEEE.
- Lei Wang, Boyi Zeng, Steve Lin, Guangyou Xu, and Heung Yeung Shum. 2004. "Automatic Extraction of Semantic Colors in Sports Video", *ICASSP '04, Montreal, Canada, May*.
- M. R. Naphade and T. S. Huang. 2001. "A probabilistic framework for semantic video indexing, filtering, and retrieval". IEEE, *Transaction On Multimedia*, 3: 141-151.
- Muriel Gastaud and Michel Barlaud. 2002. the "Video Segmentation using Active Contours on a Group of Pictures". Los Alamitos: IEEE.
- SangHyun Jang and HyungJe Cho. 1997. "The Improved Techniques of

- Automatic Scene Change Detection for Content Based Indexing on Compressed Video Data”. Seoul: Dongguk University.
- Shuqiang Jiang, Qixiang Ye, Wen Gao, and Tiejun Huan. 2001. “Wen Gao and Tiejun Huan”, “A New Method to segment Playfield and Its Applications in Match Analysis in Sports Video”. New York: ACM.
- Sohaib Khan and Mubarak Shah, 2001. “Object Based Segmentation of Video Using Color, Motion and Spatial Information”. Los Alamitos: IEEE.
- Y. Alp Aslandogan and Clement T. Yu, 1999. “Techniques and Systems for Image and Video Retrieval”. Los Alamitos: IEEE.
- Yueting Zhuang, Ruogui Xiao, and Fei Wu, 2003. “Key Issues in Video Summarization and Its Application”. New York: ACM.
- Zhibin Lei, Wu Chou, Jialin Zhong, and Chin Hui Lee. 2000. “Video Segmentation using Spatial and Temporal Statistical Analysis Method”. Los Alamitos: IEEE.
- Wei Li, Susan Gauch, John Gauch, Kok Meng Pua. “VISION: A Digital Video Library”. DL '96, Bethesda MD USA: ACM.