

## 비디오에서 객체의 시공간적 연속성과 움직임을 이용한 동적 객체추출에 관한 연구\*

박 창 민\*\*

### *A Study on the Extraction of the dynamic objects using temporal continuity and motion in the Video*

Park Changmin

#### 〈Abstract〉

Recently, it has become an important problem to extract semantic objects from videos, which are useful for improving the performance of video compression and video retrieval. In this thesis, an automatic extraction method of moving objects of interest in video is suggested. We define that an moving object of interest should be relatively large in a frame image and should occur frequently in a scene. The moving object of interest should have different motion from camera motion. Moving object of interest are determined through spatial continuity by the AMOS method and moving histogram. Through experiments with diverse scenes, we found that the proposed method extracted almost all of the objects of interest selected by the user but its precision was 69% because of over-extraction.

Key Words : Video Compression, Object Extraction, Moving Information, Segmentation, Moving Object

### I. 서론

최근 들어, 컴퓨터의 성능 향상, 고속의 네트워킹, 인터넷 사용의 급증으로 인하여 대용량의 비디오 정보를 공유하는 것은 더 이상 어려운 일이 아니다. 비디오는 텍스트와 같은 매체와는 달리 사용자가 쉽게 많은 정보를 습득할 수 있도록 해주고, 비디오로 이루어진 학습은 텍스트로 이루어진 것 보다 학습 효

율을 높게 해준다. 비디오는 교육 방송, 디지털 도서관, 전자 상거래 등과 같은 많은 분야에서 새로운 응용의 중요한 부분으로 인정되고 있으며, 점점 더 많은 응용 분야에서 활용될 것이다. 한편 게임에서 게임 캐릭터의 정서를 연구하는 분야[1]에서도 게임 캐릭터를 분석하여 정서구조를 연구하는데 캐릭터를 활용하고 있다. 뿐만 아니라 화면 분할을 효율적으로 사용하여 영상에 포함된 콘텐츠의 효율성을 높이고자 하는 연구 분야[2]에도 다양하게 응용될 수 있다.

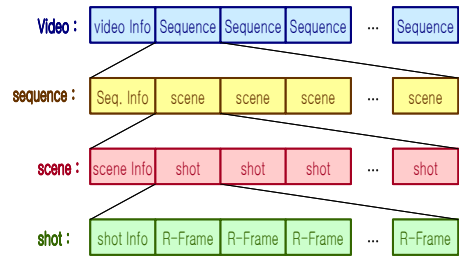
그러나 비디오 처리가 수작업에 의해 이루어지

\* 본 연구는 2016년 영산대학교 교내 연구비의 지원을 받아 수행되었음.

\*\* 영산대학교 자유전공학부 교수

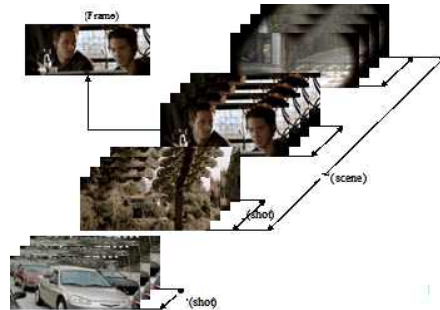
고 있어, 대량의 비디오를 처리하기 위해 많은 비용이 요구된다. 그리고 비디오를 이용한 인터넷 서비스는 아주 열악한 수준이다. 현재 가장 잘 알려진 비디오 서비스의 예는 비디오 검색이다. 현재의 비디오 검색은 비디오의 제목, 파일이름, 비디오가 존재하는 문서의 내용 등에 포함된 키워드를 이용한 방법이 주류를 이룬다. 따라서 비디오의 내용이 사용자가 원하는 정보와 다를 수 있으며, 사용자는 그 비디오를 다운로드 후 처음부터 끝까지 보아야만 그 내용을 파악하고 원하는 정보인지를 확인할 수 있다. 현재 서비스 중인 비디오 검색 사이트에서는 키워드 검색 방법으로 접근하고, 내용과는 연관성 없는 비디오 포맷 타입, 실행시간, 다운로드 유무, 프레임 개수 등의 비디오 정보만을 제공하기 때문에, 검색 결과의 정확성과 신뢰성이 떨어진다. 이러한 문제점을 해결하기 위해서 내용 기반 비디오 검색, 비디오 요약 및 browsing 등의 기술 개발이 요구된다.

그러나 비디오는 연속된 많은 영상으로 이루어진 매체로, 순차적인 특징을 가진다. 대량의 비디오를 처리하기 위해서 걸리는 시간과 공간을 생각해서라도, 가장 먼저 이루어져야 하는 비디오 처리는 순차적인 데이터를 관리하기 쉬운 단위로 변경시키고 중복을 줄이는 작업이다. 이는 비디오를 분할하는 작업을 통해서 해결할 수 있다. 비디오는 논리적으로 <그림 1>과 같은 계층적인 구조로 분할할 수 있다. 변화가 발생하지 않거나 비슷한 영상들을 합병하고 이들을 대표할 수 있는 영상을 추출함으로써 중복을 줄이고 관리하기 쉬운 형태들로 변경하여 처리시간과 공간을 줄이는 효과를 얻게 있다. 이러한 작업을 샷(shot) 검출, 대표 프레임(Representation Frame, R-Frame) 추출이라고 한다.



<그림 1> 비디오의 계층적 구조

먼저, 샷(shot)이란 <그림 2>와 같이 한 번의 카메라 녹화에 해당되는 연속적인 프레임들을 의미하고, 카메라 동작의 기본 단위이다. 따라서 샷(shot) 검출이란 연속적인 비디오에서 샷 전환이 발생한 지점을 찾는 것이다. 여기서 샷 변화는 카메라의 동작에 의한 변화를 의미한다. 대표 프레임은 샷을 대표하는 프레임을 의미한다.



<그림 2> 비디오의 shot과 scene

<그림 1>에서 샷은 한 개 이상의 대표 프레임을 가질 수 있음을 보여준다. 대표 프레임을 이용하여 비디오 처리가 이루어지므로 계산량이 줄어든다. 그러나 대량의 비디오를 처리하기 위해 걸리는 시간과 공간이라는 문제를 제외하더라도 가장 근본적인 문제는 다양한 내용이 포함되어 있어서 개인별로 같은 비디오에 대한 주관적인 해석이 서로 다를 수 있다는 것이다. 그래서 단순히 카메라의 동작에 의해 발생하는 샷 검출과는 다른 고수준의 분할 기법이 요

구된다.

한편 내용 기반 비디오 검색을 위해서는 효율적으로 비디오를 표현할 필요가 있고, 의미를 가지는 기본 단위인 객체를 정확하게 추출하고 잘 표현해 낼 필요가 있다. 비디오에서의 객체 추출은 여러 프레임간의 움직임 정보 및 칼라 정보를 이용하여 추출할 수 있다. 비디오에서 객체 추출 방법은 크게 칼라나 텍스처 정보에 기반을 두어 영역을 분할하는 영역기반 객체 추출 방법과 움직임 정보에 기반을 두어 영역을 분할하는 움직임 기반 객체 추출 방법으로 나눌 수 있다. 움직임 기반 객체 추출 방법은 움직임 정보를 이용하여 각 프레임들을 유사한 움직임을 가지는 영역으로 분할하여 객체 영역을 대략적으로 추출하고, 칼라정보를 이용해 영역들을 세밀하게 추출한다. 비디오에서 유사한 장면을 검색하기 위해서 장면의 특징을 잘 추출하고 표현하는 것이 무엇보다 중요하다. 따라서 영역 기반 객체 추출 방법[3]과 움직임 기반 객체 추출 방법[4] 등 여러 가지 객체 추출 방법들이 연구되었다. 그러나 객체 추출을 위해서 사용자가 질의 초기에 객체의 대략적인 형태를 일일이 입력해야 하는 단점들이 있다. 뿐만 아니라 [5]은 첫 프레임에 대하여 영역을 분할하고 연속되는 프레임들에 대하여 프로젝션 추적을 통하여 영역 분할을 시도 하였다. 그리고 배경 영역간 움직임 유사도를 이용하여 병합하고 나머지 전경 영역에 대하여 시간적 제약을 이용하여 객체를 추출 하였다.

본 연구에서는 움직임 정보와 칼라 정보에 기반을 두어 비디오에서의 동적객체를 추출하는 방법을 제안한다. 동적객체를 정의하고 추출하는 것은 비디오 압축과 검색에 유용하게 사용될 수 있고 사용자의 검색 의도와 검색 질의 표현 사이의 의미적인 차이를 극복하기 위한 중요한 표현 수단이 될 수 있다.

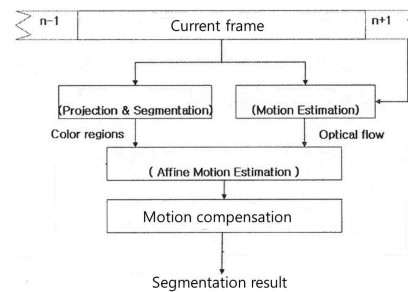
2장에서는 AMOS[6, 7] 방법의 움직임 객체 추출

결과를 이용하여 동적객체를 결정하고 추출하였다. 3장에서는 중심 영역에서 동적객체의 움직임을 판별한다. 마지막으로 4장에서는 실험을 통하여 제안된 결과들을 분석하며, 5장에서 결론을 맺는다.

## II. 동적객체추출

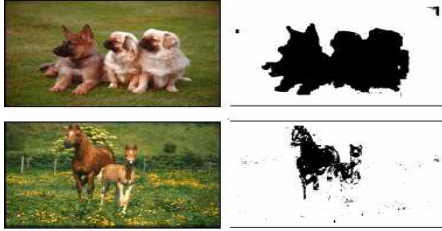
### 2.1 AMOS 에 의한 영역 분할

AMOS(Active MPEG-4 Object Segmentation)는 프로젝션을 통한 영역 분할과 움직임 추정 과정들을 통하여 유사변환 움직임을 추정하고 마지막으로 움직임 보정을 통하여 각 프레임의 영역을 분할한다 <그림 3>. 프로젝션 과정에서 이전 프레임의 영역 정보를 이용하기 때문에 영역의 추적이 가능하다. 또한 에지 검출과 양자화를 통해 공간적인 칼라 정보를 이용한다.



<그림 3> AMOS 영역분할 및 추적 다이어그램

움직임 정보는 계층적인 블록 매칭 방법을 사용하여 움직임 벡터 값들을 추출한다. 그리고 유사 변환 추정을 위하여 선형회귀(linear regression) 알고리즘을 사용하여 마지막 단계에서 로그 스텝(log(D)-steps)영역 매칭을 사용하여 보정한다.

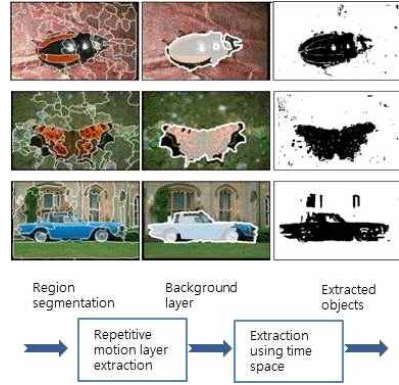


<그림 4> AMOS 영역분할 결과

<그림 4>는 AMOS 방법을 통한 한 프레임의 영역 분할 결과를 보여준다. 영역 분할 결과는 다음 프레임에 영향을 주기 때문에  $n$  번째 프레임에서 나타난 영역이  $n+1$  번째 프레임에서 또 다른 형태로 나타날 수 있다. 이렇게 영역 추적을 통해 구해진 영역들은 프레임간의 관계를 보여주므로 이후 객체 추출의 과정에서도 매우 중요한 역할을 한다.

## 2.2 시공간적 연속성을 이용한 동적객체추출

동적객체 추출을 위해, 먼저 AMOS 영역 분할 결과를 이용하여 두 단계 방법으로 움직임 객체를 구한다. 반복적인 움직임 레이어 추출 단계에서는 움직임이 유사한 영역들을 병합하고 반복적으로 배경 레이어를 추출한다. 배경 레이어는 항상 연속적인 움직임 경계를 가지기 때문에 각 레이어의 경계 움직임 벡터의 변화율을 가지고 배경 레이어를 추출한다. 배경 레이어를 제외한 영역들을 전경 레이어로 정하고 이들 중 가장 핵심적이고 관심이 있는 객체를 시간적으로 자주 나타나고 공간적으로 큰 영역들의 집합을 움직임 객체로 추출한다. <그림 5>는 동적객체를 추출하는 과정과 입력으로 주어진 영역분할 결과와 중간결과 및 최종 움직임 객체 추출 결과를 보여주고 있다.



<그림 5> 영역분할과 추적기반 추출된 동적객체의 두 단계

## III. 중심영역 움직임 판단

### 3.1 움직임 히스토그램

중심 영역에 움직임을 판단하는 것은 중심에 정비 되어 있으면서 의미가 있는 객체가 있을 가능성 여부를 판별하는 것과 같다. 중심 영역에 움직임의 분포를 알기 위해 장면 내의 모든 프레임에 대하여 각 픽셀 별로 움직임이 나타나는 정도를 히스토그램으로 구한다.

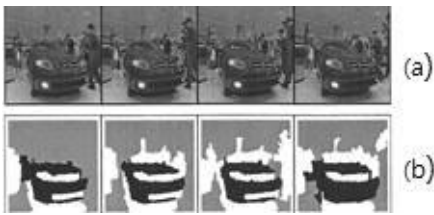
수식(1)에서  $i, j, f$  는 각각 가로좌표, 세로좌표, 프레임 번호를 나타내며,  $Output f(i, j)$ 는  $f$ 번째 프레임의  $(i, j)$  위치가 움직임 객체에 포함되면 1 값을, 아니면 0 값을 반환한다. 결국 한 장면에 대하여 2차원 히스토그램  $Hist(i, j)$ 이 생성된다.

$$Hist(i, j) = \sum_{f=1}^N Output_f(i, j),$$

$$Output_f(i, j) = \begin{cases} 1, & (i, j) \in \text{Moving object of AMOS} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

<그림 6> 은 장면의 여러 프레임 중에 9, 12, 15,

18, 번째 프레임을 보여주고 있다. (b)는 장면 내의 움직임 히스토그램을 나타내고 있다. 한 장면에 대하여 하나의 움직임 히스토그램이 생성될 수 있음을 알 수 있다. (b)에서는 (a)장면의 움직임 분포를 눈으로 쉽게 보기 위해 장면 내의 총 프레임 수에 대한 움직임 분포 히스토그램 값을 0~255의 값으로 변환하여 보여주고 있다.



<그림 6> 움직임 분포 히스토그램

### 3.2 중심영역 움직임 빈도 계산

정지 영상뿐만 아니라 비디오에서도 중심영역은 사용자의 관심 영역이다. 비디오는 움직임 정보, 음성 정보 등 많은 정보를 가지고 있기 때문에 사용자의 관심이 다양하다. 하지만 중심 영역에 움직이지 않는 객체가 있고, 움직이고 있는 객체들에 대한 관심도가 떨어진다면 중심 영역의 중요성은 더욱 커진다. 여기서 중심 영역을 기본 관심 창으로 정의하고 영상의 25% 이상을 차지하는 사각형 영역을 의미한다.

중심 영역의 움직임 빈도는 수식(2)의 CMA (Central Motion Activity)로 정의된다. 움직임 빈도는 기본 관심창 위치에 속하는  $Moving(i, j)$  값들의 합으로 나타낸다. 여기서  $Moving(i, j)$ 는 움직임 분포 히스토그램의 값이 일정 값 이상인 것은 1 값을, 그렇지 않은 것은 0 값을 가진다.  $Moving(i, j)$ 가 1 값을 갖는다는 것은 (i, j)가 여러 프레임에 걸쳐 움직임이 빈번하게 나타나는 위치라는 것을 의미한다.

$threshold_1$ 는, 장면 내 총 프레임 수의 50%가 되는 값으로 실험 하였다.

$$\begin{aligned} & \text{if } Hist(i, j) > threshold_1, \text{ } Moving(i, j) = 1 \\ & \text{otherwise } Moving(i, j) = 0 \end{aligned} \quad (2)$$

$$\begin{aligned} CMA &= \sum_{(i,j) \in DAF} Moving(i, j) \\ & \text{if } CMA < threshold_2, \text{ } No \text{ Moving Activity near Center} \end{aligned}$$

<그림 7>은 중심 영역 움직임 정도 값이 낮게 나타나는 장면 (a)와 높게 나타나는 장면 (b)를 보여주고 있다. (a)는 중심 영역에서 한 사람이 권투를 하고 있는 장면이고 (b)는 자동차 시연을 위해 모델과 사람들이 주변에서 움직이며 구경하는 장면이다. (a)(b)의 가운데 그림은 움직임 분포 히스토그램을 0~255로 변환하여 나타내고 있다. (a)(b)의 가장 오른쪽 그림은  $Moving(i, j)$  값이 1인 것은 검은색으로 0인 것은 흰색으로 보여준다. 따라서 (a)는 중심 영역의 움직임이 빈도가 높고 (b)는 움직임 빈도가 0에 가깝다는 것을 알 수 있다.



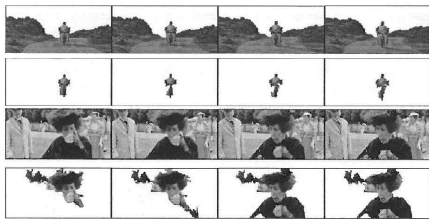
<그림 7> 중심영역 움직임 실험 결과

## IV. 실험 및 토의

본 연구에서는 비디오에서의 시공간적 연속성 및 중심 영역 움직임 정도를 측정하고 움직임 히스토그램과 빈도를 계산하여 최종적으로 동적 객체를 추출하였다. <그림 8>은 동적 객체가 추출된 결과를 나

타내고 있다. 움직임이 중앙 부분에 두드러지게 존재하는 비디오 장면들의 예를 볼 수 있다. 첫 번째 비디오는 오토바이가 멀리서 부터 가까이 다가오면서 카메라도 조금씩 먼 쪽에서부터 가까운 방향으로 움직인다. 중앙부분에 주변과 다른 움직임을 가지고 있어서 동적객체가 비교적 뚜렷하게 추출 되고 있는 것을 알 수 있다.

두 번째 비디오는 움직임이 두드러지게 이루어지지만 전반적으로 어둡고 움직임이 없는 프레임 영역에서는 부분적으로 객체로 추출하지 못하는 영역이 생기고 배경이 과잉 추출되는 영역 또한 발생하였다. 하지만 전반적으로 동적객체 부분을 잘 추출해내고 있음을 알 수 있다.



<그림 8> 동적 객체 추출 결과

<표 1, 2> 는 10편의 비디오에서 총 40 장면을 대상으로 제안한 방법을 실험한 결과이다. 표 1은 실제로 사람이 관심 객체라고 생각하는 것과 제안한 방법의 결과 추출된 객체들이 대략적으로 일치하는가? 그렇지 않은가를 비교하고 있다. True/False 는 실제로 사람이 관심객체라고 생각하는 것과 그렇지 않은 것을 의미하고 Positive/Negative 는 제안된 방법에 의해 추출된 결과를 나타낸다.

<표 1> 실제관심 객체와 제안한 방법의 비교 결과

	Positive	Negative
True	31(TP)	1(TN)
False	14(FP)	-

<표 2> 제안된 방법의 정확도 평가 결과

Recall	Precision	f-measure
0.96	0.69	0.80

## V. 결론

비디오에서 동적객체 추출은 압축과 검색의 성능을 향상 시키는 중요한 요소이다. 본 연구에서는 비디오 장면 내의 동적 객체의 히스토그램 정보와 프레임 간 핵심 객체 영역의 유사도 정보 및 움직임 정보를 이용하여 객체 영역을 병합하고 의미 있는 동적객체를 추출하는 방법을 제안 하였다. 동적객체 추출 방법은 앞으로도 중요한 과제로 남아 있다. 먼저 정확한 동적객체를 추출하기 위하여 신뢰성 있는 움직임 정보가 필요하다.

또한 카메라의 움직임에 의해 발생될 수 있는 다양한 변환을 정확하게 나타낼 수 있는 방법이 필요하다. 결론적으로 정확하면서 계산 량이 적은 동적 객체 추출 방법을 찾는 것이 향후 인터넷 및 모바일 기반 응용에도 중요한 과제이다.

## 참고문헌

- [1] 박이슬, 고일주, 박준형 “온라인 게임내 캐릭터의 정서 구조 연구,” 디지털산업정보학회 논문지, 제12권, 제3호, 2016, pp. 167-179.
- [2] 주현식, “영상미디어 형태 분류에 따른 화면 분할에 관한 연구,” 디지털산업정보학회 논문지, 제11권, 제2호, 2015, pp. 131-139.
- [3] P. Salembier and F.Marques, “Region based representations of image and video:

- segmentation tools for multimedia services,"  
IEEE Trans. Circuits syst. Video Technol., Vol.  
9, pp. 1147-1169, 1999.
- [4] Fatih Porikli, Yao Wang, "Automatic video  
object segmentation using volume growing and  
hierarchical clustering," JASP, 2004.
- [5] R. Venkatesh Babu, K. R. Ramakrishnan and S.  
H. Srinivasan, "Video Object Segmentation: A  
compressed Domain Approach," IEEE Trans.  
Circuits syst. Video Technol., Vol. 14, April,  
2004.
- [6] S. F. Chang et al, "VideoQ: An automated  
content based video search system using visual  
cues," Proc of ACM Multimedia Conference,  
Nov. 1997.
- [7] Di Zhong and shih-Fu Chang, "AMOS -An  
Active MPEG-4 Video Object Segmentation  
system," ICIP, Oct. 1998.

■ 저자소개 ■



박 창 민  
Park Changmin

2016년 12월~현재  
영산대학교 자유전공학부 교수  
2006년 2월 부산대학교 컴퓨터공학과  
(공학박사)  
1996년 2월 독일 도르트문트대학교  
컴퓨터공학과(공학석사)  
1993년 2월 독일 도르트문트대학교  
컴퓨터공학과(공학사)  
관심분야 : 영상검색, 영상이해 및 추출  
E-mail : cmpark@ysu.ac.kr

논문접수일 : 2016년 12월 7일  
수 정 일 : 2016년 12월 13일  
게재확정일 : 2016년 12월 13일