

# 양방향 멀티미디어 방송을 위한 객체 추출 및 부호화 시스템 구현

유홍연\*, 이지호\*, 홍성훈\*

\*전남대학교 전자공학과

e-mail : [keister@vip.chonnam.ac.kr](mailto:keister@vip.chonnam.ac.kr)

## Object Extractor and Encoder for Interactive Multimedia Broadcasting

Hong-Yeon Yu\*, Ji-Ho Lee\*, Sung-Hoon Hong\*

\*Dept. of Electronics Engineering, Chonnam University

### 요 약

본 논문에서는 멀티미디어 방송 및 인터넷 등에서 객체기반 멀티미디어 방송 서비스를 제공하기 위한 영상 객체 추출 및 편집 부호화 시스템을 구현 하였다. 구현된 시스템은 연속된 영상에서 객체 추출을 위해 카메라로부터 입력되는 영상에 대한 자동분할과 사용자의 조력에 의한 반자동분할 기능을 가지고 있다. 또한 추출된 객체를 기존의 저장매체에 저장되어 있는 객체 혹은 배경 영상과 합성하는 편집 기능을 제공하고, 합성된 영상을 국제표준인 MPEG-2 영상부호화 방식을 이용하여 부호화를 수행한다. 따라서 구현한 시스템은 비디오 정보 제공자가 편리하게 객체를 분할하고 편집할 수 있으며 비디오 정보를 단기간에 효율적으로 비디오 정보 사용자에게 공급할 수 있는 기능을 제공한다.

### 1. 서론

최근 디지털 방송에 있어 대화형 양방향성 서비스가 가시화되고 인터넷 방송이 활성화됨에 따라 비디오 정보 제공자가 특정 인물이나 배경 같은 객체를 새로운 객체로 대체하는 등 디지털 방송용 동영상에서 특정 객체를 제거하고 추적하는 객체단위의 처리가 중요시 되고 있다. 또한 객체기반 부호화 및 전송, 영상편집, 영상검색시스템, 감시시스템 등 다양한 객체기반 멀티미디어 응용기술을 위한 객체분할 방법이 필요하게 되었다.

이러한 객체분할은 무엇보다도 신뢰성 있고 사용자의 개입을 최소화하며 편리한 분할 방법을 요구하고, 이에 따른 많은 객체분할 방법이 제안되었다<sup>[1][2][3]</sup>. 그러나 이러한 객체분할의 목적은 연속된 영상에서 배경과 객체를 분할하는 방법만을 제시할 뿐 실제 대화형 멀티미디어 방송 등을 위한 객체기반 영상조작, 부호화 등과 같은 방법은 제시하지 않고 있다.

따라서 본 논문에서는 이러한 객체분할 방법들을

응용하여 비디오 정보 제공자가 쉽고 빠르게 사용자가 요구하는 비디오 정보를 생성하여 제공할 수 있는 통합된 시스템을 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서는 구현된 시스템의 각각의 기능을 서술하고, 3장에서는 구현된 시스템 및 자동/반자동분할에 의한 객체추출 실험결과를 보이고, 4장에서 결론 및 향후 연구를 기술한다.

### 2. 객체분할 및 편집 부호화 시스템

본 논문에서 구현한 객체분할 부호화 시스템의 구조는 그림 1과 같다. 구현한 시스템의 기능은 실시간으로 카메라를 통해 들어오는 영상 데이터를 각각의 밝기와 색상 값의 차이를 이용한 마스크를 생성하고 이 마스크를 이용한 객체분할을 수행하는 자동 분할과 사용자의 직관적인 조력과 공간 정보 및 객체추적에 의한 반자동분할을 통해 객체분할을 수행한다. 그리고 분할된 객체와 DB(data base)로 저장되어 있는 배경 혹은 객체를 사용자의 요구에 따

라 추출된 객체와 합성하고 MPEG-2로 부호화를 한다. 각각의 수행 과정은 다음과 같다.

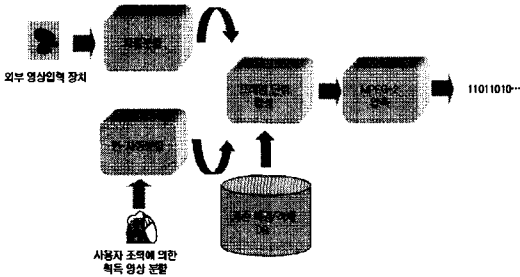


그림 1 시스템 구성도

### 2.1 자동 객체분할

제안한 자동 객체분할기에서는 객체가 포함되지 않는 배경 프레임과 새로운 객체가 나타나는 현재 프레임과의 프레임차를 매 프레임 마다 적용하여 변화되는 객체를 추출한다. 제안된 자동분할기의 구조는 그림 2와 같다.

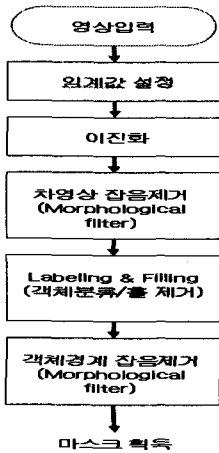


그림 2 자동 분할을 위한 마스크 획득

우선 카메라를 통해 실시간으로 입력되는 배경영상을 버퍼에 저장한다. 이 후 새로운 객체가 출현한 영상과 버퍼에 저장된 배경영상과의 밝기와 색상 정보에 대한 차이값을 구하여, 이 차이값 중 가장 큰 값에 비중을 둔 단일 차영상을 생성한다. 이렇게 생성된 차영상에서 배경과 움직임이 나타난 영역을 구분짓는 두 영역의 경계값을 기준으로 임계값을 설정하고 임계값을 기준으로 원영상을 분할하여 초기 마

스크인 이진화 영상을 생성한다. 생성된 이진화 영상에 대해서 모폴로지 닫힘과 열림 필터의 연속연산을 적용하여 작은 오류 영역이나 경계 주변에 존재하는 미세한 오류영역을 제거하여 영상을 단순화 한다. 이 모폴로지 필터 과정을 거친 영상에 레이블링 과정을 통해서 개별적인 객체 인식 과정을 수행한다. 레이블링 과정 후 생성된 객체 내의 홀 성분을 필링 과정을 수행함으로써 객체 내의 오류를 제거한다. 이 후 다시 모폴로지 열림과 닫힘 필터의 연속연산을 적용하여 객체 경계의 가지성분의 잡음을 제거하여 최종 마스크를 획득한다. 이렇게 획득된 객체를 표현하는 마스크 정보를 이용하여 현재영상에서 배경과 객체를 분할한다. 이는 적은 연산량으로 효율적인 객체분할을 수행하고 실시간으로 입력되는 영상에 신뢰성 있는 객체추출을 수행한다.

### 2.2 반자동 객체분할

구현한 반자동 분할기는 연속된 영상에서 첫 번째 영상의 의미있는 객체를 추출 하는 Intra-frame 분할과 초기 추출된 객체의 움직임 정보를 추적하여 분할하는 Inter-frame 분할 방법으로 구성되어 있다 [1]. 그림 3은 반자동 객체분할의 처리과정을 나타낸 것이다.

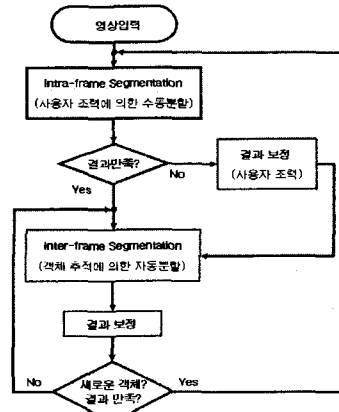


그림 3 반자동 분할기 구조

Intra-frame 분할은 입력된 영상의 첫 번째 영상에서 의미있는 객체에 대해 그림 4와 같이 사용자의 조력에 의하여 객체의 경계를 분할한다.

이 분할된 경계를 공간영역의 밝기와 색상정보에 대해 Morphological filter를 이용한 영상 단순화를 수행하여 잡음 성분을 없애고 각각의 밝기 및 색

상 값 중 가장 큰 그래디언트 값을 갖는 영상을 구한다. 그리고 이 그래디언트 영상을 이용하여 사용자의 조력에 의해 획득된 객체의 경계와 실제 객체와의 불확실한 경계영역에 대해 Immersion Simulation에 의한 Watershed<sup>[4]</sup> 알고리즘을 적용하여 초기 객체를 추출한다. 또한 잘못 추출된 경계정보에 대해서는 사용자의 직관적인 조력에 의해 잘못 설정된 경계선을 마우스 드래그로 수정함으로써 정밀한 초기 객체를 생성 할 수 있다.



그림 4 사용자 조력에 의한 경계 분할

Inter-frame 분할에서는 이전 프레임에서 획득한 객체 경계정보 마스크를 블록매칭(block matching) 알고리즘을 적용하여 현재 프레임에서 이동된 객체 경계의 움직임 벡터를 구하고, 추정된 움직임 벡터로부터 움직임 파라미터를 산출한다. 또한 추정된 움직임 벡터에 메디안 필터를 적용하여 잡음 성분을 제거하고 보다 정밀한 움직임 정보를 획득한다.

이렇게 이전 프레임으로부터 획득된 객체 경계정보와 산출된 움직임 파라미터를 이용하여 이전 프레임에 속한 객체를 현재 프레임에 투영함으로써 연속된 프레임 상에서 자동으로 객체를 추출한다. 그리고 정확한 객체의 경계를 추출하기 위하여 투영된 경계의 불확실한 영역을 산출하고 이 불확실한 영역을 watershed 알고리즘을 사용한 영역 확장을 통해 정밀한 객체경계를 추출하는 보정 기능을 제공한다.

### 2.3 합성기 및 부호화기

합성기에서는 자동분할 및 반자동분할에 의해 획득된 객체를 기존의 DB로 저장되어 있는 배경 혹은 객체 영상과 합성 기능을 수행한다. 그림 5의 오른쪽 상단의 창에 기존의 DB로 저장되어 있는 목록을 보여 준다. 따라서 비디오 정보 제공자는 수요자가 원하는 배경을 쉽게 선택하고 분할된 객체와 합성을 효율적으로 수행할 수 있다.

부호화기에서는 합성된 영상을 국제표준인 MPEG-2 영상부호화 방식을 이용하여 윈도우 미디어 플레이어에서 재생할 수 있는 압축 비트스트림으

로 만들어준다.

### 3. 구현결과 및 실험

구현된 시스템은 Windows XP 환경에서 Visual C++ 6.0을 이용하여 구현하였고, 카메라의 입력을 위해 Windows에서 제공하는 통합 비디오 인터페이스인 VFW(Video for Window)을 이용하였다.

그림 5는 구현된 시스템의 사용자 환경을 보여 주고 있다. 왼쪽 상단의 창은 자동/반자동 분할을 위한 입력 영상을 나타내고, 중앙의 창은 현재 처리되는 영상의 결과를 보여준다. 또한 반자동 분할기에서 사용자 개입을 위한 작업 창을 제공한다. 그리고 오른쪽 상단의 영역은 기존의 DB로 저장되어 있는 배경 및 객체 영상의 프레임 수, 사이즈 정보를 사용자에게 제공하고 합성을 원하는 배경을 선택할 수 있는 편리한 기능을 제공한다.

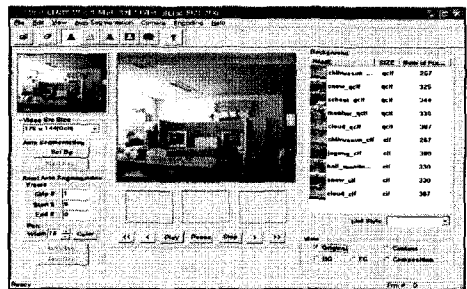


그림 5 구현된 시스템의 사용자 환경

그리고 중앙의 하단에 있는 세 개의 창은 왼쪽부터 각각 추출된 마스크, 객체, 배경 영상을 실시간으로 디스플레이 해줌으로써 비디오 정보 제공자가 객체 추출 상황을 실시간으로 모니터링 할 수 있는 편리함을 제공한다.

그림 6은 실시간으로 입력되는 영상으로부터 자동분할에 의해 추출된 객체 영상을 보여준다.

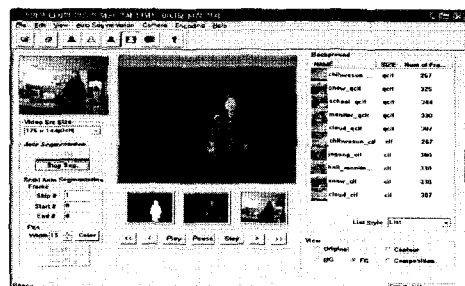


그림 6 자동분할 결과

그림 7은 반자동분할에 의해 추출된 객체 영상들과 실제 비디오 영상을 추출된 객체와 합성한 결과 영상을 보여 주고 있다. 객체 추출을 위한 입력영상으로는 Akiyo, 352x288 사이즈의 영상을 사용하고 비디오 영상을 밝기와 색상정보만을 갖는 영상으로 변환하여 초당 30 프레임으로 실험한 결과이다.

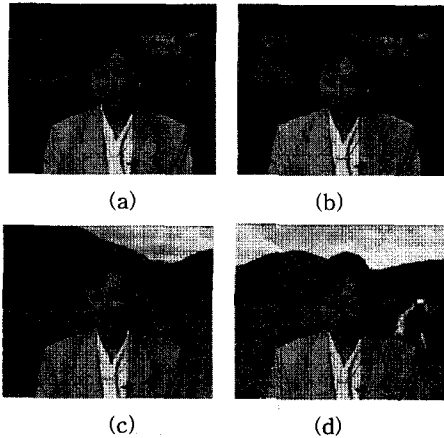


그림 7 (a) 사용자 개입에 의한 초기 추출 영상, (b) 200th 영상에 대한 객체 추적 영상, (c) 초기 영상에 대한 합성 결과, (d) 200th 영상에 대한 합성 결과

그림 8은 합성된 영상을 MPEG-2 부호화 방식을 이용하여 1.5M로 부호화하는 환경과 부호화된 영상을 윈도우 미디어 플레이어로 재생한 결과를 보여 주고 있다.

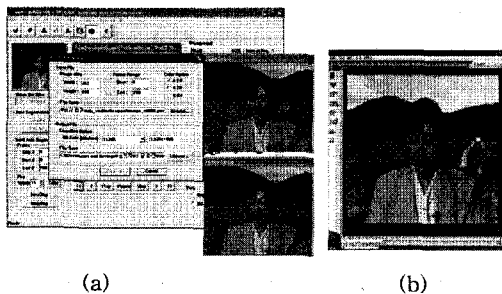


그림 8 (a) MPEG-2 부호화 환경, (b) 윈도우 미디어플레이어 재생 화면

#### 4. 결론 및 향후 연구

구현된 시스템은 여러 가지의 실험 결과를 보듯이 비디오 정보 제공자가 쉽고 간편하게 객체를 분할하고 합성 할 수 있으며 국제표준인 MPEG-2 영상부호화 방식을 지원 하는 영상 재생 틀에서 간단

히 재생된 영상을 볼 수 있다. 또한 구현된 자동분할기는 적은 계산량으로 신뢰성 있는 객체를 실시간으로 추출 할 수 있었으며, 반자동 분할기 역시 정밀한 객체를 추출 할 수 있었다. 그리고 인덱스 화되어있는 저장 영상으로부터 사용자가 간편하게 추출된 영상과 합성을 할 수 있고 간단하게 MPEG-2로 부호화를 수행 할 수 있다.

향후 연구로는 본 논문에서 제안한 합성된 영상을 프레임 단위의 MPEG-2로 부호화하는 것을 객체기반 부호화인 MPEG-4를 이용하여 객체단위로 부호화를 수행하여 높은 압축율을 제공하고, 실사영상과 3D 객체를 합성할 수 있도록 MPEG-4 시스템을 적용하여 인터넷 방송, 영상회의, 감시시스템에 적용할 수 있는 시스템을 개발하여 비디오 정보 공급자 및 사용자에게 보다 효율적인 객체기반 부호화 시스템을 제공할 것이다.

#### 참고문헌

- [1] J.G.Choi, M.C.Kim, M.H.Lee, and C.D.Ahn, " A User-Assisted Segmentation Method for Video Object Plane Generation," ITC-CSCC'98, July, pp. 7-10, 1998.
- [2] Chuang Gu, Ming-Chieh Lee, "Semiautomatic segmentation and tracking of semantic video objects," IEEE Trans. on CSVT, vol. 8, Issue 5, pp. 572 -584, Sep. 1998.
- [3] 이인재, 김규현, "효율적인 반자동 움직임 객체 분할 알고리즘 연구," 통신 및 신호처리소사이터티 춘계학술발표, pp. 841-844, 2002.
- [4] Luc Vincent and Pierre Soille, "Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 13, no. 6, Jun. 1991.
- [5] ISO/IEC 13818-2, "Recommendation H.262: Generic coding of moving pictures and associated audio information: Video," ISO/IEC JTC1/SC29/WG11, Nov. 1994.