

〈主 題〉

내용기반 영상압축을 위한 영상분할

원 치 선

(동국대 전자공학과)

□ 차 례 □

I. 서 론

II. MPEG-4를 위한 영상분할의 조건

III. 객체분할 알고리즘

IV. 결 론

I. 서 론

정지 및 동영상 데이터를 압축하기 위해 지금까지 사용된 방법들은 주로 픽셀 혹은 고정된 크기의 블록 단위로 압축하는 기법들이었다. 이와 같은 픽셀 혹은 블록단위의 파형기반 영상압축 기술들은 소위 제 1세대 영상압축기술로 분류되며, 영상통신을 위해 지금까지 발표된 JPEG, MPEG-1,2, H.261, H.263 등은 바로 이 제 1세대 영상 압축기술들을 바탕으로 제정되었다. 그러나, 고효율의 압축기술이 요구되는 8-64kbps 급의 초저속 영상통신용이나 멀티미디어 응용을 위해서는 주어진 영상의 내용을 인식하여 영상압축에 반영하는 내용기반(content-based) 영상압축 기술이 필요하다.

제2세대 영상압축기법으로 분류되는 내용기반 영상 압축기법은 주어진 영상을 의미 있는 동질의(homogeneous) 영역(즉, 객체 등)으로 분할한 다음 분할된 객체의 컬러, 움직임, 그리고 형태(shape) 정보가 압축된다. 이와 같이 의미 있는 객체단위(object-based)로 압축이 실행되므로 최소한 다음의 두 가지 중요한 특성이 기대된다. 첫째, 고압축율에서 제 1세대 영상압축 기법의 최대 단점인 블록효과(blocking effect)를 방지할 수 있으며, 분할된 객체 내부의 컬러 특성에 맞추어 적절한 영상부호화 기술을 선택적으로 적용할 수 있으므로 인간 시각특성

(human visual system)을 고려한 고효율 영상압축 기법을 실현할 수 있다. 둘째, 분할된 객체들이 각각 독립적으로 부호화되므로 멀티미디어 응용에서 요구되는 객체단위의 다양한 처리가 가능하다. 예를 들어, 객체기반(object-based) 영상압축 기법으로 압축된 비트열로부터 특정객체의 발생과 소멸까지의 움직임 정보를 추적할 수 있으며, 이 3차원 객체를 추출하여 변형하거나 편집에 활용할 수 있다. 실제로 이와 같은 기능을 갖춘 객체기반 영상압축 기술을 수용하는 동영상 처리 표준안이 구체화되고 있다. 즉, 1998년 11월 국제표준(IS)으로 확정될 소위 "비디오/오디오 객체의 부호화(coding of audiovisual objects)"를 표방하는 MPEG-4 [1][2]와 "멀티미디어 내용 서술 인터페이스(multimedia content description interface)"를 표방하며 멀티미디어 내용의 검색을 규정하기 위해 최근에 표준화 움직임이 일고 있는 MPEG-7 [3]이 그것이다.

객체기반 영상압축 기법들은 그림 1과 같이 해석단계(analysis phase)와 합성단계(synthesis phase)의 두 단계로 이루어져 있다[4]. MPEG-1,2와 같은 제 1세대 영상압축 기법에서는 주어진 영상을 고정된 크기의 블록으로 나누어 부호화하므로 그림 1의 해석 단계를 거치지 않는다. 그러나 MPEG-4와 같이 객체단위로 압축하는 경우에는 객체분할을 중심으로 하는 해석단계를 필수적으로 거치야 하며, 합성단계와는

독립적으로 수행될 수 있다. 그러나 높은 압축 효율을 달성하기 위해서는 해석단계와 합성단계에 강력한 상호작용이 요구된다[5]. MPEG-4 표준안은 합성단계는 규정하고 있으나, 해석단계는 표준화하지 않고 있다. 따라서 사용자들의 선택에 의한 고유의 해석단계(즉, 영상분할) 알고리즘을 사용할 수 있으며, 해석단계에서 분할된 영상의 결과에 따라 MPEG-4와 MPEG-7의 성능에 커다란 영향을 미치게 된다. 또한, MPEG-4가 1998년 국제표준으로 확정된다 해도 합성단계의 입력으로 제공될 분할된 객체(VOP: Video Object Plane)를 실시간 내에 만들어 줄 영상분할 알고리즘이 존재하지 않는 한 MPEG-4에 적용될 동영상은 극히 제한적일 수 밖에 없을 것이다. 예를 들어, 애니메이션이나 그래픽 혹은 블루 스크린으로 제작된 영상물 등 자연영상을 제외한 특수 제작된 영상물에 국한되어 사용될 수 밖에 없을 것이다.

도록 하며, 영상내의 객체들을 독립적으로 부호화/복호화할 수 있도록 해주는 내용기반 기능이다. 이 내용기반 기능이 가능하도록 하기 위해 오디오와 비디오의 내용에 따른 객체를 정의할 필요가 있다. 즉, MPEG-4의 내용기반 기능을 부여하기 위해 주어진 동영상은 독립적으로 액세스될 수 있는 오디오/비디오 객체(AVO)로 구성되어야 한다. 이 중 움직이는 실체 즉, VO (Video Object)는 VOP(Video Object Plane)라 불리는 2차원 공간 속의 객체들의 시간 연속적인 집합으로 구성되어 있다. 예를 들어, 그림 2는 여성의 전면객체 (foreground object)와 뒷 배경객체 (background object)의 두 개의 VOP로 이루어져 있다고 볼 수 있다. 한편, MPEG-4의 또 다른 목적은 압축효율의 향상에 있으며, 이 목적을 달성하기 위해 각 VOP는 압축효율을 높일 수 있는 동질 (homogeneity)의 데이터로 이루어져야 한다. 즉, 그림

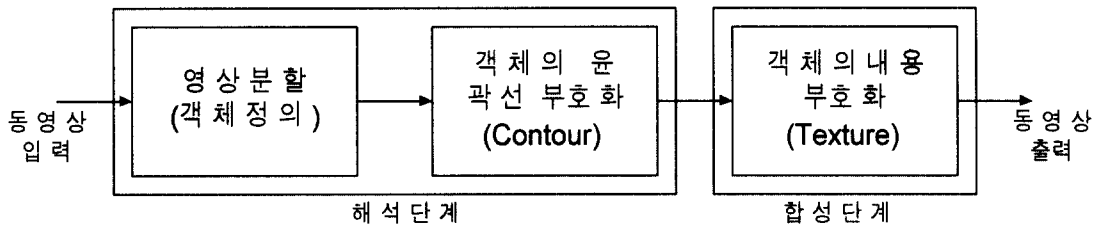


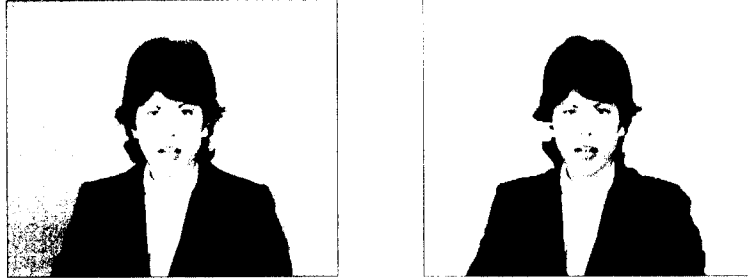
그림 1. 객체기반 영상압축 블록도

본 고에서는 내용기반 영상압축 기술 중 핵심기술로 떠오르고 있는 영상분할(image segmentation) 기법에 대해 지금까지 제안된 방법을 중심으로 살펴보고 앞으로의 발전 방향을 예측해 본다.

II. MPEG-4를 위한 영상분할의 조건

동영상의 내용 중 흥미 있는 부분에 대해 독립적 및 선택적으로 부호화하고 복호화하는 소위 “내용기반 단계적 포함성 (content-based scalability)”이 MPEG-4에서 요구하는 중요하고 특징적인 기능 중의 하나다. 이 기능은 동영상의 내용을 압축된 데이터 비트 스트림으로부터 직접 대화식으로 처리할 수 있

2의 여성객체도 얼굴이나 몸체의 밝기의 세기나 질감의 특성에 따라 그림 3에서와 같이 더욱 작은 여러 개의 sub-VOP로 나뉘어 질 수도 있다. 비록 밝기의 세기가 서로 다르더라도 움직임이 같은 인접된 영역을 모아 구성한 (즉, 움직임 분할된) 객체도 움직임 보상 에러를 줄일 수 있으므로 압축효율을 높일 수 있는 객체로 볼 수 있다. 결국 VOP를 얻기 위한 해석단계의 영상분할은 지향하는 목적에 따라 다음의 두 가지 경우로 분류될 수 있다[1].



a) 원 영상

b) 분할된 얼굴 객체

그림 2. 표현을 위한 영상분할의 예

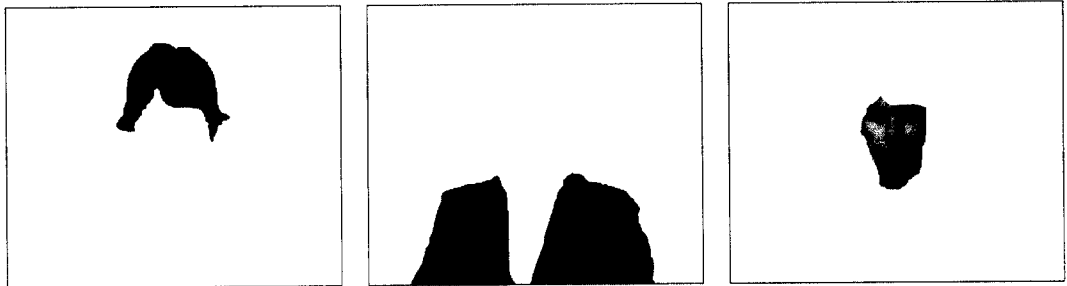


그림 3. 부호화를 위한 영상분할의 예

- 표현을 위한 영상분할 : 객체의 액세스(access)와 운용(manipulation)의 기본 단위가 될 수 있도록 의미적(semantic) 동질성 기준에 의한 영상분할
- 부호화를 위한 영상분할 : 압축효율을 높일 수 있는 동질성 기준에 의한 영상분할

물론, 궁극적으로 위의 두 가지 목적을 동시에 만족하는 영상분할이 내용기반 기능과 압축효율의 측면에서 제2세대 영상압축 기법의 목표가 되겠으나, 응용분야에 따라 두 가지 목적 중에 어느 한 쪽에 더욱 큰 비중을 둘 수 있다. 위의 두 가지 영상분할 기준 중에 첫 번째 “표현을 위한 영상분할” 기준은 두 번

째의 “부호화를 위한 영상분할” 기준에 비해 실제 영상에 적용하여 자동적으로 해당 VOP를 얻기가 상대적으로 더욱 어렵다. 따라서 VOP를 얻기 위한 난이도 혹은 동영상의 생성환경 등의 조건에 따라 다음과 같이 여러 단계의 영상분할이 가능하다.

- 자동적 영상분할 (automatic segmentation 혹은 unsupervised segmentation)
- 반자동적 영상분할 (semi-automatic segmentation with human guidance)
- 인위적 영상분할 (hand-drawn segmentation)
- 촬영기법에 의한 영상분할 (studio segmentation)

자동적 영상분할은 주어진 영상분할 기준에 따라 컴퓨터가 자동적으로 VOP를 생성하는 것으로 자연 영상에 대해 궁극적으로 지향하는 영상분할 방법이다. 그러나, 현존하는 기술로는 실시간 내에 의미적 및 동질적 기준을 충분히 만족하는 VOP를 추출할 수 있는 알고리즘은 아직 보고되지 않고 있다. 차선책으로 실시간 실행조건을 배제하면 반자동적 영상분할로도 비교적 빠르고 손쉽게 분할된 객체를 얻을 수 있다. 예를 들어, 영상내 원하는 VOP의 윤곽선을 인간의 개입으로 개략적으로 그리고 (혹은 약속된 표시를 하고), 이 결과를 초기 분할 조건으로 하여 컴퓨터가 나머지 더욱 정밀한 윤곽선을 찾도록 할 수 있다[6]. 반자동적 영상분할과 인위적 영상분할은 모두 인간의 개입을 요구한다. 그러나, 반자동적 영상분할에서 인간과 컴퓨터가 영상분할 과정의 절반씩을 담당한다면 인위적 영상분할은 전적으로 인간의 도움으로 영상분할을 얻는다는 차이점이 있다. 따라서 인위적 영상분할의 처리 시간이 상대적으로 더 많이 걸린다. 촬영기법에 의한 영상분할은 블루스크린 등의 특수 촬영기법을 이용하여 물체와 배경 등을 동영상 제작 당시 미리 분리하는 것으로 별도의 분할 과정은 필요하지 않다.

위의 영상분할 분류 중에 자동적 영상분할은 실시간 처리를 추구하는 반면, 나머지 분할방법들은 실시간 처리가 필수적인 요구조건이라고 볼 수 없다. 영상의 실시간 처리 조건의 유무에 따라 영상분할 기술을 다음과 같이 나눌 수 있다[7].

- 비실시간 분할 (off-line segmentation)
- 실시간 분할 (real-time segmentation)

비실시간 분할은 저장용용을 위한 객체의 추출 (즉, 멀티미디어 저작도구 등)에 응용될 수 있으며, 반자동적 영상분할에서와 같이 인간의 개입이 허용된다. 물론, 인간의 도움이 적고 자동화가 많이 될수록 더욱 바람직하다. 실시간 분할은 대화식 응용에서와 같이 실시간 처리를 요하는 곳에 적용되므로 분할 알고리즘의 시간복잡도가 중요한 성능평가 요소가 된다. 실시간 분할에서 추출된 객체가 후에 다시 쓰일 가능성이 희박하므로 정확한 형태(shape) 추출은 불필요하다. 반면, 비실시간 분할의 경우 분할된 객체가 저장되어 후에 다시 쓰일 수 있으므로 비교적 정확한 객체분할이 요구된다.

III. 객체분할 알고리즘

MPEG-4와 MPEG-7의 고유 기능을 충분히 활용하기 위해서는 동영상 내의 3차원 객체를 추출해 낼 수 있어야 한다. 즉, 밝기의 세기(graylevel), 움직임(motion), 질감(texture), 원근(depth), 그리고 의미적 동질성(semantic)에 의해 분할된 결과들을 잘 융합하여 3차원 객체를 정의해야 한다. 본 장에서는 객체분할을 위해 사용된 이들 동질성의 기준을 살펴보고 각 동질성 기준을 적용하여 얻은 영역들의 융합에 대해서도 알아본다.

3.1 밝기 세기의 동질성에 의한 영상분할

밝기의 세기(graylevel)를 동질성의 기준으로 하여 영역을 자동분할하는 방법으로 RSST (Recursive Shortest Spanning Tree)[8], 피라미드식 영역성장[9], 수학적 형태학(mathematical morphology)[10]-[14], 그리고 클러스터링(clustering)[15] 등을 이용하는 기법이 있다. 효율적인 동영상 압축의 견지에서 윤곽선의 복잡도를 줄이면서 객체의 실제 형태(shape)에 얼마나 가까운가의 척도로 비교한 결과, RSST 기법이 가장 좋은 성능을 내고 피라미드식 영역성장이 그 뒤를 이었다[16]. 또한 계산 복잡도의 측면에서 역시 RSST가 가장 빠르고, 그 다음은 수학적 형태학에 의한 방법이었다. 수학적 형태학에 의한 방법은 동영상 압축을 위한 영상분할에 많이 시도되고 꾸준히 개발되어 온 기술이다. 이 방법으로 얻은 영상분할의 결과는 대개 작은 여러 개의 영역들로 이루어져 있으므로 그대로 부호화하는 경우 많은 부호화 비용이 든다. (물론, 구조요소 (structuring element) 선택이나 마커추출의 과정에서 영역의 크기를 일정 크기 이상으로 제한할 수 있다.) 그러나, 윤곽선의 복잡도를 줄이는 후처리 기법을 거치면 RSST와 비교될 만한 성능을 낼 수 있을 것으로 기대된다[16]. 본 장에서는 특히 수학적 형태학의 연산자를 적용한 영상분할 기법에 대해 좀 더 자세히 살펴보기로 한다.

영상분할 기반 영상압축 문제에 수학적 형태학이 자주 활용되는 것은 수학적 형태학을 이용한 방법이 다음의 두 가지 특성을 갖고 있기 때문이다[10]. 첫째로 형태학적 필터는 윤곽선 정보를 유지하면서 영역의 내부를 평탄하게 (단순하게) 만들므로써 영상

분할을 손쉽게 달성할 수 있도록 한다. 즉, 형태학적 연산자는 영역의 윤곽선을 날카롭게 유지하면서 영역 내 밝기의 세기를 평활화(smoothing)하는 상반된 요구를 잘 만족시킨다. 두 번째는 영역의 윤곽선을 워터셰드(watershed) 알고리즘으로 효율적으로 얻을 수 있기 때문이다. 이와 같은 장점들에 고무되어 참고문헌 [10]에서는 공간적 영역에서의 영상분할을 형태적 연산자에 의해 얻었으며, 참고문헌 [11]에서는 이 결과를 시간적 영역을 고려한 비디오 영상물에 적용하였다. 형태학적 연산자를 이용한 계층적 영상분할은 각 계층단계에서 그림 4와 같은 블록으로 구성 되어 있다[11].

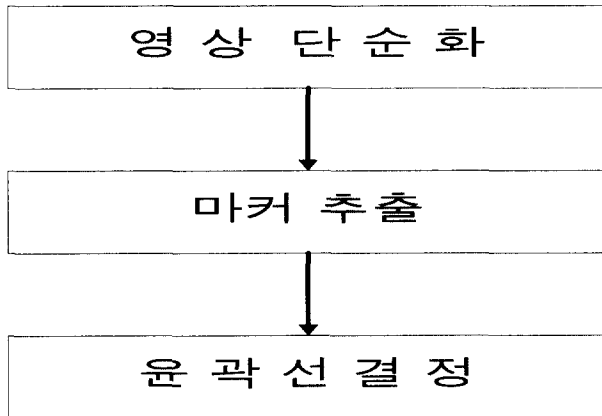


그림 4. 형태학적 연산자를 이용한 영상분할의 구성도

3.1.1 단순화 과정

단순화 과정은 주어진 영상이 쉽게 분할될 수 있도록 영역내 밝기의 세기를 평활화(smoothing) 시키는 과정이다. 이때, 영상 데이터의 단순화를 위한 공간적 전처리 필터가 사용된다. 이 전처리 필터는 윤곽선을 날카롭게 살리면서 영역내의 밝기 값을 일정하게 유지시키는 역할을 하므로써 영상분할을 손쉽게 달성할 수 있도록 한다. 이와 같은 목적을 위해 사용되는 필터로 복원적 열림닫힘(open-close by reconstruction)에 의한 형태학적 연산자를 들 수 있다.

복원적 필터(filters by reconstruction)는 "연결 연산자(connected operator)"라 불리는 좀 더 큰 필터집단에 속하는 것으로 일정한 신호(평탄영역) 끼리만 상호작용하는 특성을 갖고 있다 [12][13]. 따라서, 이

필터는 기존의 다른 필터 (예를 들어, 선형필터, 미디언 필터, 혹은 형태학적 열림-닫힘 필터)와는 달리 신호 (혹은 영상)의 다른 주파수 성분을 제거하지 않는다. 결과적으로 영상내 윤곽선의 날카로움을 유지하면서 영역내부를 평활화시키는 바람직한 특성을 갖는다. 즉, 연결연산자는 픽셀들에게 주파수 선택적인 병합(merge)이 아닌 평탄영역(flat zone) 선택적인 병합을 이룰 수 있도록 한다. 또한, 연결 연산자들에 의해 생긴 평탄영역들은 피라미드 구조에 의해 서로 포함관계로 연결될 수 있다. 이때, 계층적 단계의 레벨 상승에 따라 새로운 영역이 생기지는 않고, 기존의 평탄영역이 그대로 있거나 통합되므로 이에 따른 경계선이 유지 또는 제거된다. 따라서 연결연산자를 이용한 영상분할은 계층적 구조에 유리하며 형태학을 이용한 객체분할이 계층적 구조를 갖는 것도 해상도의 단계적 포함성(resolution scalability)의 목적 이외에 연결연산자의 바로 이런 특성을 고려한 결과다 [14].

3.1.2 마커 추출

마커(marker)는 동질성(homogeneity)을 갖는 영역이 존재함을 나타내는 라벨(label)로써 연속된 동질의 픽셀들에게 붙여진다. 즉, 마커 추출은 단순화 과정에서 확인된 영역 내부에 고유의 라벨을 붙이는 과정이다. 단순화 과정이 에지정보를 유지하면서 평탄영역을 출력한다는 사실로 부터 마커추출 방법은 평탄영역을 추출하여 얻는 방법과 밝기의 대조(contrast)를 이용하는 두 가지 방법이 있다[11]. 평탄영역을 이용할 때는 최소 크기 이상의 연결된 평탄영역들에 대해 큐(queue) 구조를 이용하여 각각 유일한 라벨값을 할당한다. 마커로 부터 영역의 정확한 경계를 확인할 수는 없지만 최소한 영역의 내부를 확인할 수는 있다.

3.1.3 윤곽선 결정

영역의 내부를 나타내는 마커가 결정된 다음에는 마커와 마커 사이의 아직 소속이 확정되지 않은 픽셀들의 영역을 결정하므로써 영역의 윤곽선을 나타낼 수 있다. 이를 위해 워터셰드(watershed) 알고리즘이 사용될 수 있다. 특히, 영역성장에 기반을 둔 워터셰드 알고리즘은 영역의 내부를 차지하는 마커로부터 시작하여 영역내 모든 공간을 점유할 때 까지 영역을 확장하는 알고리즘이다[4][11]. 이 워터셰드 알고리

즘을 효율적으로 수행하기 위해 계층적 큐(hierarchical queue)가 필요하다. 즉, 우선 순위에 따라 여러 개의 큐를 구성한 다음, 큐에서 뽑아낼 첫 번째 픽셀은 가장 높은 우선 순위의 큐에서 가장 먼저 들어간 픽셀이 되며, 이와 같은 순서에 따라 우선 순위가 높은 쪽에서 낮은 쪽으로 그리고 큐에 들어온 순서대로 처리된다. 이 계층적 큐를 이용하여 영역성장 기반 워터셰드 알고리즘의 처리 과정은 다음과 같다[4].

- 큐의 초기화 (initialization) : 라벨값이 확정된 영역내부의 모든 픽셀들의 위치정보를 가장 높은 우선 순위의 큐에 넣는다.
- 영역성장 (region growing) : 큐의 우선 순위대로 뽑은 픽셀이 어떤 영역에 아직 소속되지 않았다면 현 픽셀과 각 이웃하는 영역들과의 거리를 계산한 뒤 가장 짧은 거리를 나타내는 영역에 해당 픽셀을 할당한다. 만약 현재의 픽셀과 이웃하는 픽셀 중에 아직 영역에 할당되지 않은 픽셀이 있으면 계산된 거리에 해당되는 우선 순위로 큐에 이들 픽셀을 넣는다. 따라서 큐에 들어 있는 픽셀들은 적어도 하나의 이웃영역을 갖게 된다.

위 영역성장 과정을 큐의 우선 순위대로 반복하여 큐가 빌 때까지 계속하면 마커와 마커사이의 할당되지 않은 픽셀들이 모두 하나의 영역에 소속되어 윤곽선이 확실한 영역분할 결과를 얻는다. 이 알고리즘의 주요 파라미터는 역시 우선 순위를 결정할 때 사용한 거리의 측정기준이 될 것이다. 특히, 영상압축을 위한 영상분할의 경우 윤곽선의 복잡도(즉, 비선형성)가 부호화의 비용을 높일 수 있으므로 윤곽선 결정의 정확도와 윤곽선의 복잡도 사이에 상보관계를 제어할 수 있는 거리 함수를 사용한다. 예를 들어, 거리 함수를 픽셀의 밝기값 y 와 영역 R 의 평균값의 차이의 절대치와 영역의 윤곽선 변화량에 비례하는 값과의 가중치의 합으로 다음과 같이 정의할 수 있다[4].

$$d(y, R) = \alpha \left(y - \frac{\sum_{t \in R} t}{|R|} \right) + (1 - \alpha) \Delta \theta R \quad (1)$$

여기서 $|R|$ 는 영역 R 내의 모든 픽셀의 수를 나타내고, α 는 윤곽선의 정확도와 복잡도 사이의 가중치를 나타내는 파라미터로 $\alpha=0$ 인 경우는 거리의 기준을 모두 윤곽선 복잡도(즉, 선형화)에 두고, $\alpha=1$ 인 경우는 윤곽선의 복잡도는 고려하지 않고 윤곽선의 정확도만 반영하는 함수가 된다. 만약 영역의 3차원 트래킹(tracking)을 고려하면 식 (1)의 거리함수에 영역의 찌그러짐(deformation)의 측정기준을 포함시킬 수 있다[5].

3.2 움직임의 동질성에 의한 영상분할

움직임 벡터는 줌(zoom)이나 패닝(panning)등 카메라의 움직임에 기인한 전체적 움직임(global motion)과 화면내 물체의 변위에 기인한 국부적 움직임(local motion)으로 나눌 수 있다[17]. 또한 움직임의 원근(또는 깊이)에 따라 전면 움직임(forward motion)과 후면 움직임(backward motion)으로 나눌 수 있다[18]. 전면 움직임과 후면 움직임을 구별하는 것과 전체적인 움직임을 계산하는 것은 상호 연관성이 있다. 즉, 전체적 움직임 파라미터를 후면에서 계산하고, 이에 따라 추출된 후면을 따로 처리하는 스프라이트(sprite)를 영상압축에 활용하여 동영상 부호화의 성능향상에 기여할 수 있다. 이런 점을 고려하여 배경과 전면 영상을 구분하는 2단계 전체적/국부적 움직임 예측방법이 제안되었다[18]. 전체적 움직임을 예측하는 방법으로 정합방법(matching approach), 차분방법(differential approach), 국부적 움직임 필드에 의한 회귀방법(regression) 등이 있는데 그중 정합방법이 비디오 부호화를 위한 전체적 움직임 예측으로 가장 적절한 것으로 평가된다[18].

여러 가능한 3차원적인 움직임이 2차원 평면에 투영되면 똑 같이 하나의 움직임으로 나타나므로 연속적인 두 프레임에 의한 2-D 움직임 추정은 유일한 해가 존재하지 않는 잘 정의되지 않은(ill-posed) 문제이다. 따라서 움직임 추정을 위해 추가적인 모델(혹은 사전지식, 가정)이 필요하다. 지금까지 제안된 여러 움직임 모델들은 파라메트릭 움직임 모델(parametric motion model)과 논파라메트릭 움직임 모델(non-parametric motion model)로 나눌 수 있다. 파라메트릭 움직임 모델은 기저함수들의 선형조합과 그들의 계수(파라미터)로 움직임을 나타낸다. 즉, 평면

객체의 고정(rigid) 3-D 움직임의 경우 6개의 파라미터를 갖는 어파인 (affine) 모델로 나타낼 수 있고, 좀더 복잡한 경우 8개의 파라미터나[19] 12개의 모델 파라미터를 갖는 2차 평면으로[20] 확장될 수 있다. 따라서 주어진 영역이 같은 움직임을 갖는다고 가정할 때 그 영역을 대표하는 움직임을 이들 소수의 모델 파라미터로 표현할 수 있다. 이때, 모델의 파라미터는 프레임 차이가 최소가 되도록 하는 최소자승법 (least square)으로 구하거나, 수정된 블록정합(block matching), 혹은 Hough 변환 기법을[21] 사용할 수 있다.

3-D 고정 움직임 등과 같이 단순한 움직임에 대해서는 파라메트릭 모델이 잘 적용되나 복잡한 움직임을 갖는 경우 픽셀당 최소한 1개의 독립적인 움직임 벡터를 사용하는 논파라메트릭 모델이 더 유용하다. 논파라메트릭 모델은 각 픽셀당 움직임 벡터에 공간적 연속성의 제한조건 (smoothness constraint)을 부여하여 움직임 추정에 사용한다. 이때, 공간적 연속성의 조건을 결정적(deterministic)으로 사용하느냐 아니면 통계확률적(stochastic)으로 사용하느냐에 따라 광학적흐름방정식(optical flow equation) 이나 베이시안(Bayesian) 방법들이 사용된다. 예를 들어, 베이시안 방법은 공간적 연속성의 사전지식(a priori knowledge)를 GRF(Gibbs random field) 모델의 클릭(clique) 파라미터에 반영한다. 즉, 움직임 보상에러(Y)를 정적(stationary) 백색 가우시안 잡음(P(Y|X))으로 가정하고 움직임 벡터(X)의 공간적 연속성을 GRF (P(X))에 반영하므로써 MAP(Maximum a posteriori) 추정 $P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$ 에 의해 움직임을 예측할 수 있다[22][23].

파라메트릭 움직임 모델은 영역내 모든 픽셀들이 동일 움직임을 갖는다고 가정한다. 따라서 움직임 예측을 위한 대상영역이 동일 움직임을 갖지 않는 경우 에러가 발생하며, 이 문제를 해결하기 위해 돌출인자(outlier)에 강한 예측기 (robust estimator)를 사용할 수 있다[24]. 어파인 움직임 모델의 경우 움직임 매개 변수를 최소자승법에 의해 구할 때 정합(matching) 알고리즘과 차분(differential) 알고리즘을 사용할 수 있다. 정합 알고리즘은 블록정합 알고리즘의 확장된 개념으로 배경영역 전체에 대한 정합을 실행한다. 밝기 신호의 정보 없이 움직임 에러를 최소화하여 움직임 모델 파라미터를 계산한다. 이에 반해

움직임 정보와 밝기 정보를 통합한 통합유사성척도 (joint similarity measure)를 활용하여 개체분할의 성능을 향상시킨 시도도 있다 [25]. 차분방법은 밝기신호의 모델을 이용하여 1차 또는 2차 근사화를 위한 최소화 문제를 통해 전체적 움직임 (global motion)을 6개의 어파인 모델 파라미터로 표현하였다.

움직임 정보를 영상분할에 활용하는 방법으로 영상분할과 움직임 예측을 동시에 적용하는 방법과 공간적인 영상분할과 시간적인 움직임 분할을 따로 얻은 다음 이들 결과를 융합하는 방법이 있다. 참고문헌 [26]에서는 MRF (Markov Random Field) 모델과 베이시안 기준을 이용하여 시공간적 영상분할과 움직임 예측을 동시에 달성하는 알고리즘을 제안하였으며 참고문헌 [17]에서는 공간적인 영상분할과 움직임 정보를 이용하여 분할과 움직임 예측을 연속적으로 개선하는 알고리즘을 제안하여 공간적인 영역에서 얻은 영상분할 결과를 움직임 정보에 따라 영역을 분리 또는 흡수하는 기법을 제안하였다. 즉, 지금까지 제안된 움직임을 고려한 분할기반 영상압축 알고리즘들은 공간영역의 분할정보와 움직임 정보 사이의 상대적 중요도에 따라 다음과 같이 나눌 수 있다. 우선, 공간영역의 동질성 기준에 의한 영역분할 [27][28][29]과 움직임 정보가 주요 동질성의 기준으로 사용되는 경우[26][30]로 나눌 수 있다. 물론 공간영역과 움직임의 두 가지 동질성 기준을 동시에 적용할 수도 있다 [5][31][32]. 공간적인 영역의 동질성 기준은 영역분할의 정확도를 높이며 새로 발생하는 영역들을 잘 정의할 수 있고, 움직임 정보의 동질성 기준을 이용하는 경우 공간적인 영역들 중에 같은 움직임의 영역들을 통합하여 움직임 보상을 하는 단위로 사용하므로써 영역의 개수를 제한할 수 있으며 따라서 부호화 비용을 조절할 수 있다[5].

3.3 질감 동질성에 의한 영상분할

영상압축을 위한 영상분할에서 질감(texture) 혹은 칼라(color)라 함은 흔히 영역내의 밝기의 세기(graylevel)를 나타낸다. 그러나 영상처리 및 컴퓨터 비전에서 말하는 질감(texture) 영역은 질감요소(texture element)의 반복적 복사에 의해 형성된 패턴 및 거칠음을 말한다. 본 절에서 말하는 질감은 영상처리 및 컴퓨터 비전에서 말하는 후자를 의미하는 것으로 한다. 질감영역의 내부는 보통 밝기의 세기의

차가 커서 밝기의 세기의 동질성에 의한 영상분할의 결과로 대개 작은 여러 개의 영역으로 나누어 진다. 따라서 윤곽선이 복잡해지고 이에 따른 부호화 비용이 증가한다. 예를 들어, 그림 5-a)에서 모자 위 깃털 영역은 밝고 어두운 깃털 고유의 질감으로 형성된 부분으로 그림 5-b)와 같이 하나의 영역으로 분할하는 것이 부호화 효율과 의미 있는 객체의 정의에 더 도움이 된다. 지금까지 영상압축을 위한 영상분할 기준으로 사용된 동질성 기준은 밝기의 세기 값이나 칼라들에 의한 것이 보통이었으며 질감특성에 의한 동질성은 특별하게 취급되지 않았다. 질감특성을 영상분할에 반영하기 위해서는 질감요소의 주기성을 고려하여 픽셀단위 보다 블록단위로 영상을 분할하는 것이 더 바람직하다. 질감영역을 포함하기 위해 블록단위로 GRF 모델을 정의하고 MAP 예측 기법에 의해 블록단위 객체분할을 얻는 기법이 제안되었다[33]. 그림 5-b)는 참고문헌 [33]에서 제안된 방법에 의해 분할된 결과로 깃털과 같은 질감영역을 하나의 영역으로 분할하므로 밝기의 세기만에 의한 분할결과 보다 분할된 영역의 수가 작아 윤곽선 부호화에 따른 압축비용을 낮출 수 있고 분할된 객체에 의미("깃털")를 부여 할 수도 있다.

3.4 원근 동질성에 의한 영상분할

영상내 객체 상호간의 원근(depth) 정보는 같은 방

향의 움직임 객체들을 전방움직임 객체와 후방움직임 객체로 분리해 주는 역할을 하며, 이들 움직임 객체의 증척시 실질적인 원근에 의해 정확한 윤곽선을 결정해 준다. 이를 위해 최근에 카메라 쌍의 불균형 측정(binocular disparity estimation)에 의한 원근정보를 바탕으로 MAP 예측기준에 의한 영상분할이 제안되었다[34].

3.5 의미적 동질성에 의한 영상분할

밝기의 세기나 질감등의 동질성 기준에 의해 분할된 작은 영역들은 의미 있는 하나의 객체(예, 사람, 집, 차 등)로 인식되기 위해 시간축상의 움직임이나 원근정보를 이용할 수 있고, 경우에 따라 의미적 정보(semantic)를 이용할 수 있다. 비디오 전화의 경우와 같이 사람의 얼굴이 주 객체로 되어 있는 동영상에 대해 객체를 구성하는 얼굴의 각 영역(즉, 머리, 얼굴, 목, 눈, 어깨 등)들 사이의 상호 의미적 관계를 이용하여 얼굴객체를 정의할 수 있다. 예를 들어, 객체내의 영역들 사이에 평균값과 상호분산(covariance) 매트릭스 값을 알 수 있다면(트레이닝 데이터로부터 최대가능성(maximum likelihood) 예측으로 결정 가능), 이 값을 이용하여 MAP 예측기준 등을 적용하여 영역들을 융합(fusion) 할 수 있다[35].



a) 원 영상

b) 깃털영역 분할 결과

그림 5. 질감을 고려한 영상분할

3.6 영역의 융합

최종적인 객체를 정의하기 위해 밝기의 세기, 움직임, 원근, 및 의미적 정보에 의해 결정된 영역들을 융합(fusion)할 필요가 있다. 물론, 앞에서 살펴 본 동질성 기준이 모두 사용되지 않고 부분적인 결과를 융합하여 객체를 정의 할 수도 있다. 예를 들어, 밝기세기의 동질성에 의한 공간적인 분할 결과에 움직임의 동질성에 의한 분할 결과를 융합하여 3차원 객체를 정의할 수 있다.

최근에는 밝기의 세기와 움직임 정보 이외에 원근 및 의미적 정보를 모두 포함하는 계층적 융합 알고리즘이 제안되었다[35]. 즉, 그림 6에서와 같이 일단 작은 평탄영역들을 통합하면서 밝기세기의 동질성에 의한 영상분할 결과를 얻는다. 이들 영역들 중에 움직임이 같은 영역들은 일단 객체의 영역에 포함될 수 있으나, 움직임이 같은 영역들 중에서도 특히 같은 원근(깊이)을 갖는 영역들 만이 객체를 구성하는 영역이 된다. 그림 6에는 의미적 정보 (semantic) 분할 결과의 융합 과정이 포함되어 있지 않으나 경우에 따라 밝기의 세기 및/혹은 움직임 동질성에 의해 분할된 영역들에 대해 직접 의미적 정보를 반영하여 최종적인 객체분할을 얻을 수 있다.

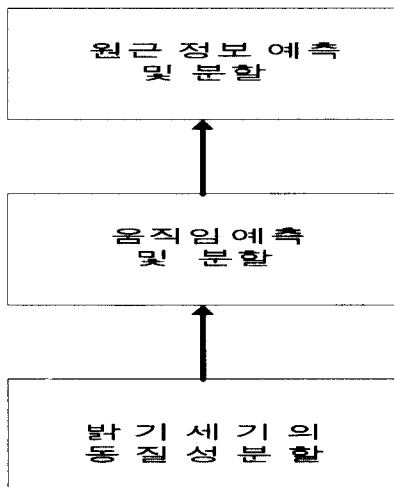


그림 6. 계층적 영역 융합

IV. 결 론

MPEG-4는 내용기반 영상압축 표준으로 의미 있는 객체의 정의가 필수적이다. MPEG-4의 표준안에는 어떻게 객체를 분할하느냐에 대해 규정하고 있지 않으므로 앞으로 MPEG-4의 구현을 위해 사용자 고유의 객체분할 알고리즘이 자유롭게 적용될 수 있다. 반대로, 고유의 객체분할 알고리즘을 갖고 있지 않는 사용자는 MPEG-4의 압축효율과 다양한 기능을 절대로 활용할 수 없다. 결론적으로 MPEG-4 및 MPEG-7의 멀티미디어 기능을 얼마나 충분히 활용할 수 있는가와 얼마나 높은 압축 효율을 달성할 수 있는가는 어떤 객체분할 알고리즘을 사용하느냐에 전적으로 달려 있다고 해도 과언은 아니다.

최근 객체분할 알고리즘의 개발 방향은 다음과 같이 요약될 수 있다.

- 밝기세기와 움직임 정보뿐만 아니라 질감, 원근, 및 의미적 정보를 활용한 영역분할 및 이들 결과의 융합방법 개발
- 제안된 객체분할 알고리즘(수학적 형태학 방법 등)의 실시간 처리를 위한 최적화 및 병렬처리 알고리즘의 개발
- 주어진 응용분야 및 조건에 따라 해당 최적 객체분할 알고리즘 개발 (즉, 자동, 반자동, 실시간, 비실시간 등의 조건에 최적인 분할 알고리즘의 독립적 개발)

[참 고 문 헌]

[1] F. Pereira, "MPEG-4: a new challenge for the representation of audio-visual information", Keynote Speech at Picture Coding Symposium'96, pp.7-16, March 1996.

[2] T. Sikora, "The MPEG-4 video standard verification model", IEEE Tran. on Circuits and Systems for Video Technology, Vol.7, No.2, pp.19-31, Feb. 1997.

[3] R. Koenen, "Overview of MPEG-7 goals and objectives", Proc. of WIAMIS'97, pp.111-116, Belgium, June 24-25, 1997.

[4] P. Salembier, et. al., "Morphological operators for image and video compression", IEEE Tran. on Image Processing, Vol.5, No.6, pp.881-898, June 1996.

- [5] P. Salembier, et. al., "Segmentation-based video coding system allowing the manipulation of objects", IEEE Tran. on Circuits and Systems for Video Technology, Vol.7, No.2, pp.60-74, Feb. 1997.
- [6] N.E. O'Connor, N. Brady, and S. Marlow, "Supervised image segmentation using EM-based estimation of mixture density parameters", Proceedings of WIAMIS'97, pp.27-32, Belgium, June 24-25, 1997.
- [7] R. Koenen, "Requirements for segmenting content to be coded using MPEG-4", MoMuSys/WG1-0020, Bergen, 25 April 1997.
- [8] O.J. Morris, M.J. Lee and A.G. Constantinides, "Graph theory for image analysis: an approach based on the shortest spanning tree," IEE Proceedings, Vol. 1333, pp.146-152, April 1986.
- [9] P.J Burt, T.H. Hong, and A. Rosenfeld, "Segmentation and estimation of image region properties through co-operative hierarchical computation," IEEE Tran. System, Man, and Cybernetics, Vol. 11, No. 12, pp.802-809, Dec. 1981.
- [10] P. Salembier, "Morphological multiscale segmentation for image coding", Signal Processing, vol.38, pp.359-386, Sept. 1990.
- [11] P. Salembier and M. Pardas, "Hierarchical morphological segmentation for image sequence coding," IEEE Tran. on Image Processing, Vol.13, No.5, pp.639-651, Sep. 1994.
- [12] J. Serra and P. Salembier, "Connecter operators and pyramids", in Proc. SPIE Image Algebra Math. Morphology, San Diego, Vol.2030, pp.65-70, 1993.
- [13] P. Salembier and J. Serra, "Flat zones filtering, connected operators, and filters by reconstructions," IEEE Tran. on Image Processing, Vol.4, No.8, pp.1153-1160, August 1995.
- [14] M. Pardas, P. Salembier, "3D morphological segmentation and motion estimation for image sequences," Signal Processing, Vol.38, pp.31-43, 1994.
- [15] P.J. Mulroy, "Spatial and temporal image segmentation by feature clustering," Int. Workshop on Coding Techniques for Very Low Rate Video, VLBV'95, Tokyo, Nov. 1995.
- [16] P.J. Mulroy, "Video content extraction : review of current automatic segmentation algorithms," Proc. of WIAMIS'97, pp.45-50, June 1997.
- [17] F. Defaux et. al., "Spatial-temporal segmentation based on motion and static segmentation," Proc. of ICIP-95, Vol. I, pp.306-309, 1995.
- [18] F. Moscheni, F. Defaux, and M. Kunt, "A new two-stage global/local motion estimation based on a background/foreground segmentation," Proc. of ICASSP-95, Vol.4, pp.2216-2264, 1995.
- [19] R.Y. Tsai and T.S.Huang, "Estimating three-dimensional motion parameters of a rigid planar patch," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-29, pp.1147-1152, Dec. 1981.
- [20] N. Diehl, "Object-oriented motion estimation and segmentation in image sequences," Signal Processing: Image Communication, Vol.3, pp.23-56, Feb. 1991.
- [21] A. Amitay and D. Malah, "Global-motion estimation in image sequences of 3-D scenes for coding applications," Signal Processing: Image Communication, Vol. 6, pp.507-520, 1995.
- [22] J. Konrad and E. Dubois, "Bayesian estimation of motion vector field," IEEE Trans. Pattern Anal. Machine Intell., Vol.14, pp.910-927, Sep. 1992.
- [23] C. Stiller, "Object-based estimation of dense motion fields," IEEE Trans. Image Processing, Vol.6, No.2, pp.234-250, Feb. 1997.
- [24] P. Anandan, et. al., "Hierarchical model-based motion estimation," In M.I.Sezan and R.L. Lagendijk, editors, Motion Analysis and Image Sequence Processing, pp.1-22. Kluwer Academic Publishers, 1993.

[25] J.G. Choi, S.-W. Lee, and S.-D. Kim, "Spatio-temporal video segmentation using a joint similarity measure," IEEE Tran. on Circuits and Systems for Video Technology, Vol.7, No.2, pp.279-288, April 1997.

[26] P. Bouthemy and E. Francois, "Motion segmentation and qualitative dynamic scene analysis from an image sequence," International Journal of Computer Vision, Vol.10, No.2, pp.157-182, 1993.

[27] P. Willemin, T. Reed, and M. Kunt, "Image sequence coding by split and merge", IEEE Tran. Communication, Vol.39, No.12, pp.1845-1855, Dec. 1991.

[28] F. Marques, V. Vera, and A. Gasull, "A hierarchical image sequence model for segmentation: application to object-based segmentation coding," Proc. of SPIE VCIP-94, Chicage, pp.554-563, Oct. 1994.

[29] P. Salembier, et.al., "Region-based video coding using mathematical morphology," Proc. of IEEE, Vol.83, No.6, pp.843-857, June 1995.

[30] H.G. Musman, M.Hotter, and J. Osterman, "Object-oriented analysis-synthesis coding of moving images," Signal Processing: Image Communication, Vol.1, No.2, pp.117-138, Oct. 1989.

[31] C. Gu and M. Kunt, "Very low-bit rate video coding using multi-criterion segmentation," Proc. of ICIP-94, Vol.II, pp.418-422, Texas USA, Nov. 1994.

[32] J. Benois, L. Wu, and D. Barba, "Joint contour-based and motion-based image sequences segmentation for TV image coding at low bit rate," Proc of VCIP-94, pp.1074-1085, Sep. 1994.

[33] C.S. Won, "Extracting edge and homogeneous regions using block-based MAP segmentation," Proc. of WIAMIS' 97, pp.167-172, June 1997.

[34] E. Francois and B. Chupeau, "Depth-based segmentation," Trans. on Circuits and Systems for Video Technology, Vol.7, No.1, pp.237-239, Feb. 1997.

[35] L. Garrido, et. al, "A hierarchical technique for image sequence analysis," Proc. of WIAMIS' 97, pp.13-20, June 1997.



원 치 선

- 1982년 2월 : 고려대학교 전자공학과 (학사)
- 1986년 2월 : Univ. of Massachusetts/Amherst (석사)
- 1990년 2월 : Univ. of Massachusetts/Amherst (박사)
- 1989년 11월 ~ 92년 8월 : 금성사 가전연구소
선임연구원(현 LG전자 멀티미디어 연구소)
- 1992년 9월 ~ 현재 : 동국대학교 전자공학과
조교수, 부교수