

# 웨이블릿 데이터의 형태적 표현을 적용한 동영상 코딩에 관한 연구

김혜경<sup>o</sup>, 오해석

승실대학교 컴퓨터학과 멀티미디어연구실

hkkim@multi.soongsil.ac.kr, oh@computing.soongsil.ac.kr

## A Study of Video Coding Based on a Morphological Representation of Wavelet Data

Hye-Kyung Kim<sup>o</sup>, Hae-Seok Oh

Multimedia Lab., Dept. of Computer Science, Soongsil University

### 요 약

영역의 수와 윤곽선의 길이는 세그멘테이션 기반의 움직임 보상된 비디오 코딩에서 두 가지의 기본적인 제약사항이다. 이 논문에서 제안하는 코딩 스킴은 영역의 수를 축소하는 것에 초점을 맞추고, 윤곽선 코딩, 그리고 치환된 프레임 차이(DFD)의 압축에 초점을 맞춘다. 제안된 스킴의 가장 중요한 특징 중의 하나는 형태적인 필터를 기반으로 하는 spatio-temporal 단순성 알고리즘이고, 그것들과 함께 이미지는 작은 수의 영역으로 나누어질 수 있다. 이 스킴의 매우 중요한 특성은 세그멘테이션 맵 샘플링 기법으로, 그것은 윤곽선 길이를 매우 작은 복원 에러에 비해하여 약 50% 까지 줄인다. 실험적인 결과는, 높은 압축 비율에 대하여 매우 작은 코딩 에러를 보여주었다.

### 1. 서론

세그멘테이션 기반의 움직임 보상 기법은 비디오 압축을 위한 매우 전도 유망한 연구분야이다. 이런 기법들은 영상을 동질적인 영역으로 세그멘팅함에 의해 시작하고, 그밖에 영역의 움직임 평가와 보상을 수행한다. 세그멘테이션 맵과 움직임 파라미터내의 윤곽선은 수신기에 전송되어야 한다. 블록 기반의 움직임 보상과 비교하여, 세그멘테이션 기반의 움직임 보상은 두가지 잇점을 가진다. 즉, 첫 번째는 더 적은 움직임 파라미터를 산출하는 것이고, 두 번째는 어떤 블록 산출물이 재구성된 영상내에서 생성되지 않는 것이다. 그러나, 의미있는 비트의 수는 세그멘테이션 맵의 윤곽선을 인코딩하기 위해 요구된다. 윤곽선의 길이가 일반적으로 영역의 수를 가지고 증가한 이후로, 영역의 수는 기본적인 제약이다. 이런 기법들의 또다른 문제는 재구성하는 새로운 오브젝트에 대한 불가능성이고, 움직임 보상을 기초로 하는 기법들에 대한 모든 것에 관해서는, DFD(displaced frame differences)로 불리는 보상 에러로 인코딩되고 또한 전송되어야한다.

본 논문에서 표현된 코딩 스킴은 영역의 수 축소, 윤곽선 코딩, 그리고 DFD 압축에 초점을 맞춘다. 스킴의 구조는 5단계를 포함한다. 즉, 형태적인 공간적-시간적 단순성, 이미지 세그멘테이션, 움직임 보상과 움직임에 의한 영역 합병, 윤곽선 압축, 그리고 세그멘테이션에 의한 DFD 코딩이다.

관련연구는 2장에 서술하고, 3장에서는 제안 알고리즘을 기술한다. 4장에서는 제안한 알고리즘의 성능을 평가한다. 마지막 5장에서 결론과 향후연구과제에 대해 서술한다.

### 2. 관련연구

형태적인 공간적-시간적인 단순성에 대한 일반적인 알고리즘이 비디오 프로세싱을 위하여 [1]에서 표현되었다. 여기서 그것은 비디오 압축에 적용된다. 이 알고리즘은 두 부분으로 구성된다. 즉, 다단계의 공간적인 단순성(multiscale spatial simplification)과 움직임-보존의 단순성(motion-preserving simplification)이다. 전자는 인간의 시각 시스템에 대하여 지각적으로 둔감한 영상의 세부사항들을 단순화하기 위해 단지 공간적인 정보를 이용한다. 후자는 정지 오브젝트를 제거하기 위해 시간적인 정보를 활용한다.

### 3. 제안 알고리즘 소개

#### 3.1 다단계의 공간적 단순성

인간의 시각 시스템을 위하여, 높은 대조를 가진 작은 컴포넌트는 아마도 낮은 대조를 이루는 커다란 컴포넌트보다 더욱 민감할 것이다. 형태적인 필터를 사용하는 것에 대한 또 다른 문제는 구조화하는 요소의 선택에 존재한다. 구조화한 큰 요소는 일부 지각적으로 민감한 세부항목을 제거할 것이고, 반면에 구조화한 작은 요소는 큰 영상 컴포넌트를 효율적으로 단순화할 수 없다. 이 문제를 극복하기 위하여,  $B_0 \supset B_1 \supset B_2$  관계를 만족하는 세 개의 구조화한 요소  $B_0, B_1, B_2$ 를 사용하고, 쓰레스홀드 T는 다단계의 공간적인 단순성에 의존한다.

$f(r)$ 은 오리지널 이미지 순서에 대한 프레임을 표현한다고 가정하고,  $r=(x, y)$ 는 공간적인 위치를 표현하고,  $f(r)$ 은 가장 큰 구조화 요소  $B_0$ 를 open-

closing 연산을 사용하여 재구성한다.  $\Psi(f)$ 에 의해 나타난 필터링 결과에서,  $f(r)$ 에 대한 커다란 동질적인 영역은 수평적 영역으로 전송되고,  $B_0$  보다 작은 모든 영상 컴포넌트는 제거된다. 영상 컴포넌트를 단순화하기 위하여 그것은 높은 대조를 가지고,  $B_0$  보다 약간 더 작은 것이고, 단순한 나머지  $f(r) - \Psi(f)$ 는 눈에 보이는(밝은) 나머지  $G_1$  과 눈에 보이지 않는(어두운) 나머지  $G_2$  로 처음 분할된다.

$$G_1(r) = \begin{cases} [f - \Psi(f)(r)], & \text{if } [f - \Psi(f)(r)] > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$G_2(r) = \begin{cases} [\Psi(f) - f(r)], & \text{if } [\Psi(f) - f(r)] > 0 \\ 0, & \text{otherwise} \end{cases}$$

두 가지의 나머지 영상에서 높은 대비 컴포넌트는 T를 가진 쓰레스홀딩에 의해 감지된다.

$$g_i(x) = \begin{cases} G_i(r), & \text{if } G_i(r) > T \\ 0, & \text{otherwise} \end{cases} \quad \text{for } i=1,2. \quad (1)$$

$g_i(x)$ 는 재구성에 의해 opening 연산을 사용하여 단순화되는 것이고  $B_1$  안에서 구조화하는 요소처럼 가져오고,  $G_i$ 도 재구성한다.

$$C_i = \gamma^{(rec)}[\gamma B_1(g_i), G_i] \quad \text{for } i=1,2 \quad (2)$$

그리고 거기서  $\gamma B_1(g_i)$ 는  $B_1$ 과 참조 Y 아래에 X의 팽창에 의한 재구성인  $\gamma^{(rec)}(X, Y)$ 을 도구로  $g_i$ 의 opening 연산을 수행한다.  $C_1$ 과  $C_2$ 는 단순화된 영상을 향상하기 위해  $\Psi(f)$ 와 함께 (3)과 같이 표현된다. 즉,  $f_1 = \Psi(f) + C_1 - C_2$  (3)

나중에,  $C_1$ 과  $C_2$ 는 각각  $G_1$ 과  $G_2$ 로부터 빼고, 그것은 두 개의 새로운 나머지를 산출한다. 수식 (1)과 (2)는 쓰레스홀드 T보다 높고,  $B_1$ 보다 적거나  $B_2$ 보다 큰 영상 컴포넌트를 단순화하기 위해 새로운 나머지에 적용된다. 이런 단계에서, (2)에 있는  $B_1$ 은  $B_2$ 로 교체되고, (2)에 대한 새로운 결과가  $f_1$ 으로 나타난다.

실험에서는 쓰레스홀드 T를 8로 가정하고,  $B_0$ ,  $B_1$ ,  $B_2$  각각은 16\*16, 8\*8, 그리고 4\*4의 평방크기로 정한다. 다단계의 공간적 단순화는 커다란 수평적 영역을 산출하는 반면에 영상의 보다 좋은 주관적인 품질을 보존한다.

### 3.2 움직임-보존의 단순성

움직임을 보존하는 단순성은 정지 컴포넌트를 제거하는 반면에 발견하고 재저장한다. 영상 순서에서 시간 t-1와 t 지점에서  $f_1(r, t-1)$ 과  $f_1(r, t)$ 은 두 개의 성공적인 프레임들을 나타낸다고 가정하고, 그것은 다단계의 공간적인 단순성에 의해 단순화되었다. 움직이는 컴포넌트의 일부분은 어떤 노이즈가 없다면

$|f_1(r, t) - f_1(r, t-1)| > 0$ 에 의해 획득될 수 있다. 2차원 공간에서 선형 구조로 나타나는 부분은 대부분 움직이는 컴포넌트에 대한 예지에 속하는 부분이다. 이 부분으로부터, 전체 움직임 컴포넌트는 회복될 수 있으므로, 선형적인 특성을 보존하는 대신에 노이즈를 줄이기 위하여 다음의 방법을 적용한다. 즉, 구별의 기준(특징)이 되는 이미지  $|f_1(r, t) - f_1(r, t-1)|$ 는 다중 방향의 형태적인 필터에 의해 여과된다. 움직임 보존의 단순화는 움직임 컴포넌트를 잘 보존하는 반면에 정지 컴포넌트를 제거할 것이고, 그러므로 오리지널 영상에 있는 몇 개의 정지 컴포넌트가 하나의 영역으로 세그멘테이션된다.

### 3.3 이미지 분할(Image Segmentation)

이미지 세그멘테이션 알고리즘은 단순화된 영상을 세그먼트 한다. 그것은 4 단계로 구성된다. 즉, 영역 성장, 영역 정제, 영역 합병, 그리고 작은 영역 제거의 단계를 수행한다.

### 3.4 움직임에 의한 움직임 평가와 영역 합병

$f(r, t)$ 는 타임 t에서 오리지널 이미지의 순서를 나타내고  $f(r, t-1)$ 은 타임 t-1에서의 이미지 순서로 디코드된다고 하자.  $f(r, t)$ 는 n개 영역  $R_i, 0 \leq i < n$ 으로 세그먼트되었다고 가정하고, 앞절에서 정의된 알고리즘에 의해,  $f(r, t)$ 와  $f(r, t-1)$  간의 움직임 파라미터는 각 영역  $R_i$ 에 대해 [10]과 [11]에서 묘사되었던 단순한 선형적인 모델을 사용하여 제거된다.

이미지  $f(r, t)$ 는 움직임 파라미터와  $f(r, t-1)$ 로 부터 재구성될 수 있다. 그러나, 영역의 수를 줄이기 위하여, 움직임 파라미터의 한 집합에 의해 완전히 재구성될 수 있다면 두 개의 인접한 지역은 병합될 것이다.

$\rho_i = (t_x, t_y, k, \theta)$ 는 영역  $R_i$ 와 영역내 움직임 분야  $d(\rho_i)$ 에 대한 움직임 파라미터를 나타낸다. 더 나은 영역 합병을 위하여, 평균 평방 보상 에러를 계산한다.

$$e(i, j) = \frac{1}{\|R_{ij}\|} \sum_{k \in R_{ij}} (f(r, t) - I[\bar{f}(r + d(\rho_j), t-1)])^2$$

for  $j = i$  또는  $j : R_j$ 는  $R_i$ 와 인접한 상태이다. 여기서  $\|R_{ij}\|$ 는 영역  $R_i$ 내에서 구성된 픽셀의 수를 나타내고, I는 쌍일차 보간법을 나타낸다.  $e(i, j)$ 는 영역이 영역 그자체에 대한 움직임 파라미터에 의해 보상되었을 때 영역  $R_i$ 에 대한 보상 에러를 표현하고, 반면에 영역이 인접한 지역  $R_j$ 의 움직임 파라미터에 의해 보상되었을 때,  $e(i, j)$ 는 영역  $R_i$ 내의 보상 에러이다.  $e(i, j) - e(i, i)$ 의 증가가 작으면, 커다란 에러 없이 영역  $R_i$ 는  $R_j$ 로 병합될 수 있다. 그러므로, 에러 증가가 가장 작은 영역에 대한 인접한 지역의 쌍을 찾는데 적용된다. 만약 가장 작은 증가가  $i=k$ 이고  $j=l$ 일때 수행된다면, 영역  $R_k$ 는 두 개의 기준에 따라  $R_l$ 로 합병된다. 즉 1)  $e(k, l)$ 은 미리결정된 쓰레

스홀드  $T_c$  보다 적다. 또는 2) 영역의 수는 미리결정된 수  $N_r$  보다 크다. 두 번째 조건은 최종적인 세그멘테이션 맵에서 영역의 수를 줄이기 위한 방법이다. 이전 영역  $R_k$  는  $R_l$  로 병합되었고,  $R_k$  와  $R_l$  과 관련된 에러는 다른 영역과 병합하기 위하여 갱신된다. 움직임에 의한 영역을 병합한 후에, 최종적인 세그멘테이션 맵에 존재하는 모든 영역에 대한 움직임 파라미터는 추정되고, 적응적으로 8 비트에 의해 양자화 되었고, 수신기로 전송된다.

3.4 윤곽선 압축

매우 단순한 샘플링 기법을 윤곽선 길이를 줄이기 위해 수정된 체인코드 기법을 세그멘테이션 맵에 적용하는 방법을 사용한다.  $S(x, y)$  는 세그멘테이션 맵을 나타낸다고 가정하자. 그것은 레이블 모드를 따라 3\*3 블록 위로 서브샘플되고, 3\*3 블록에 대한 서브샘플링 결과는 블록 이내에서 가장 유력한 레이블이다. 샘플된 세그멘테이션 맵내의 윤곽선은 체인 코드 기법에 의해 끝이 인코딩되고, 수신기로 전송된다. 수신기에서,  $S(x, y)$  는 서브샘플된 세그멘테이션 맵을 확대함에 의해 회복된다. 이런 샘플링 절차는 윤곽선 길이를 대략 50%로 줄인다.

3.5 세그멘테이션에 의한 DFD 코딩

DFD 압축에 대한 기법은 DFD의 낮은 상관 때문에 아마도 변환 기반의 기법보다 더욱 효율적이다. 세그멘테이션 기반의 코딩 기법은 적응적인 산술적 코딩에 의한 모든 픽셀의 값을 인코딩한다. 일반적으로, DFD 이미지는 일부 선형적인 구조를 구성하고 제로 값의 배경 위에 나타나는 많은 작은 지점들로 구성한다. 압축률을 높이기 위하여, DFD 영상내에서 작은 크기와 낮은 에너지 지점은 무시한다. 낮은 에너지 지점은 쓰레스홀드  $T$  지점에서 쓰레스홀딩  $D(r)$ 에 의해 제거된다.

$$d_1(r) = \begin{cases} 1, & D(r) > T \\ 0, & \text{otherwise} \end{cases} \text{ and } d_2(r) = \begin{cases} 1, & D(r) < -T \\ 0, & \text{otherwise} \end{cases}$$

그러므로 작은 크기의 지점은 다중방향의 형태적인 필터에 의해 제거된다.

예를들면,  $d_1(r)$  은 다음과 같다.

$$d_3(r) = \sqrt[r]{B_i} [d_1(r)]$$

$$d_4(r) = r^{(rec)} [d_3(r), d_1(r)]$$

여기서  $B_i$  는 길이 5와 중심  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  각각에 대한 요소를 선형적으로 구조화하는 것이다.

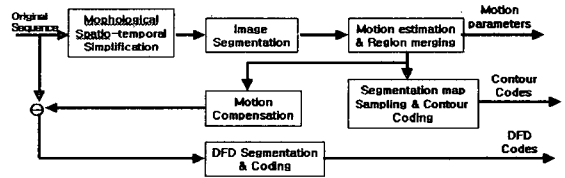
$d_4(r)$  내에서 "1"의 영역안에 아마도 일부 작은 구역이 있을 것이다. 이 구역을 제거하기 위하여,

$d_4(r)$ 은 크기 3\*3 의 요소로 구조화하는 closing 연산을 수행한다. 동일한 처리가  $d_2(r)$  에 적용된다.

두가지 결과가 곱하여 단 하나의 세그멘테이션 맵으로 결합된다. 세그멘테이션 맵 안의 각 영역은 그것의 그레이레벨 방법에 의해 근사화된다. 이것은  $D(r)$ 의 첫 번째 복원을 형성한다. 그후에, 회복 에러 이미지는 위

에서 언급된 절차에 의해 반복적으로 세그먼트되고, 결과로 생긴 세그멘테이션 맵은 어떤 새로운 영역이 획득될 수 없을 때까지 처음 세그멘테이션 맵 위에 덧붙여진다. 이런 반복적인 절차는 큰 크기에 대한 점도 아니고 선형적인 구조도 아닌 것을 보증하고 최종적인 회복 에러 영상에서 높은 에너지가 있음을 보증한다.

본 논문에서는 영역의 수를 줄이기 위하여 두가지 기준을 이용한다. 첫째, 전체 영상 지역의 0.04%보다 작은 모든 영역은 그것의 가장 유사한 인접 영역으로 병합한다. 두 번째 방법은 그들의 그레이레벨 방법에 대한 차이가 쓰레스홀드  $T$ 보다 적거나 영역의 수가 미리결정된 수  $N_r$  보다 큰 것이라면, 2개의 인접한 지역은 한 영역으로 합병된다.



(그림 1) 제안된 부호화의 전체 구조도

4. 성능평가

각 실험영상은 3절에서 설명한 영역 확장법을 사용하여 균일한 영역으로 분할되었다. 유사성의 척도로 움직임 정보와 밝기값의 연속성을 사용하였다.

	PSNR (dB)	Compression ratio	Bit rate (b/p)	Motion para.	Contours	DFD
Foreman	30.12	117.20	0.0682	15.72	45.50	38.77
Miss America	33.80	209.70	0.0380	16.27	40.37	43.33
Claire	35.68	197.99	0.0404	15.62	22.89	61.49
기존의 방법	34.01	156.30	0.0512	12.81	55.72	31.47

(표1) 평균 압축률과 PSNR

5. 결론 및 향후 연구 과제

본 논문에서는 공간적-시간적인 단순화 알고리즘이 세그멘테이션 맵에서 영역의 수를 줄이는 데 효율적이라는 것을 볼 수 있다. 단순화된 이미지 세그멘테이션과 DFD 세그멘테이션에 대하여 제안된 알고리즘은 비용이 안들고 정확하다. 윤곽선 코딩 기법은 매우 간단하고 작은 보상 에러의 증가로 전체 비트율을 줄일 수 있다. 제안된 코딩 스킴의 성능은 최적의 양자화법, 움직임 파라미터에 대한 통계상의 코딩 기법에 의해 더욱더 향상될 수 있을 것이다.

6. 참고 문헌

[1] D. Wang and J. Serra, "Morphological spatio-temporal simplification for video image segmentation," to appear in Signal Processing: Image Communication.

[2] D. Wang and C. Labit, "Morphological spatio-temporal simplification for video image segmentation," to appear in signal Processing: Image Communication.