
크기, 움직임 및 왜곡 정보에 의한 목표비트 분배

지석상*, 황재정**, 이문호*

Distribution of Target Bits based on Size, Motion and Distortion

Seok-Sang Chee, Jae-Jeong Hwang and Moon-Ho Lee

이 논문은 2000년도 군산대학교 두뇌한국21사업에 의하여 지원되었음.

요 약

이 논문에서는 객체의 크기, 움직임의 크기, 부호화 왜곡을 기준으로 다중객체간 가용비트를 효과적으로 분배하는 기법을 제안하였다. 세 변수를 이용하여 부호화하는 개념은 MPEG-4에서 다중객체를 부호화하는 데에 이용되었다. 그러나 변수들을 시각적인 관점에서 분석하여 이용하는 연구가 미흡하여 시각적으로 중요하지 않은 객체에 많은 비트가 할당되어 화질이 저하되는 결과를 초래하였다. 따라서 이 논문에서는 객체의 중요성을 세 가지 변수로 표현하고 시각적으로 비교하였다. 크기 및 움직임 정보는 입력 영상에서 미리 주어지고 왜곡량에 따라 비트를 분배하였다.

Abstract

An efficient bit rate distribution technique that distributes available bits for multiple objects based on motion vector magnitude, size of object shape, and coding distortion is presented. This coding concept using the three parameters was exploited in MPEG-4 multiple object coding. But the scheme is likely to produce poor results such as allocating more bits to less important objects and degrading picture quality, due to the lack of analysis and research in view of human visual aspect. In this paper importance of each object is represented by the three parameters and visually analyzed. Target bits are distributed according to coding distortion using the pre-assigned shape and motion information.

* 전북대학교 전자정보공학부

** 군산대학교 전자정보공학부

접수일자 : 2000. 9. 9.

I. 서 론

비트율 분배 문제가 여러 논문에서 다루어졌는데, 매크로블록 레벨에서 Pickering과 Arnold는 각각적으로 효율적인 율제어를 제안하였다[1]. 각각적 마스크인자가 제안되어 블록을 분류하는데, 마스크 인자는 공간적 변화량, 활동도, 움직임에 의해 결정된다. 다른 연구[2],[3]에서는 분산, 콘트라스트, 크기 등을 측정하여 관심영역을 결정하고 부호화에 의해 얻어지는 화질을 예측하였다. 또한 얼굴 모델을 이용하여 영상전화 장면에서 영역에 따라 다른 공간 및 시간 계위를 계산하는 기법도 제안되었다[4]. 통계적 다중화(StatMUX) 문제에서[5], 채널용량이 효율적으로 이용되도록 양자화변수 조정이 제안되었다. 이 방법은 프로그램간 통계적 변화량에 크게 의존하며 각각에 대한 균일한 화질을 얻고자 하였다.

이러한 배경하에 ISO/IEC MPEG-4 표준은 진보된 영상정보 개념을 지원하며 광범한 비트율과 응용을 위한 범용 영상음향 부호화를 시도하였다[6]. MPEG-4에서 영상정보는 비디오객체(VO: video object) 개념으로 구성되며 임의형상의 시변 영상으로 표현되어 개별적으로 다루어지고 하나의 장면을 구성하기 위해 비슷한 객체와 조합되기도 한다. VO에 연관된 정보는 비디오객체레이어(VOL: video object layer) 집합으로 표현된다. 각 레이어에는 주어진 시공간 해상도에 관한 정보를 포함하여 계위 전송 및 저장용 가능케 한다. 하나의 비디오 시퀀스가 다중의 VO로 분할되어 처리하는 경우 n 개의 VOL이 얻어진다. 각 VOL은 비디오객체평면(VOP: video object plane)의 시퀀스로 간주되어서 주어진 시간적 사상에 관한 정보를 포함하고 전통적인 비디오 프레임에 해당한다. VOP내에 형상과 텍스처는 서로 다른 것으로 부호화되며 예측된다. 형상이 먼저 부호화되고 텍스처는 예측-변환부호화의 변형이며 영상정보를 이용하여 부호화된다[7]. 하나의 비디오 시퀀스는 다중의 VOL로 나누어지고 VOL은 다시 해당 VOP로 이루어지고 하나의 VOP는 부호화 단위인 매크로블록으로 나누어진다.

다중객체(MO: multiple object) 부호화에서 서로 다른 객체간 비트율 분배는 사용자 또는 응용에 따

라 객체에 부여하는 값에 따라 할당된다. 객체는 예를 들어 가능한 최적의 전역화질을 얻는 데 관심이 있도록 하여 같은 중요도로 간주할 수 있다. 그러나 일반적으로 가장 적절한 객체가 최상의 화질로 부호화될 것을 사용자는 원한다. 예를 들어 영상회의 시스템에서 화자나 감시 시스템에서 침입자 등이다. 결국 적절성에 따른 부호화의 질이야 말로 사용자 대화성을 증가시키는 방법이다. 율제어는 응용 및 통신 환경이 부여한 제한사항을 반영하며 동작을 최적화하기 위해 부호기에 규준을 가하는 것이며 MO의 경우에 비트할당 기능과 관련하여 부호기의 중요한 요소가 된다[8]. 이 논문은 신호원의 시각적 모델링과 비용척도를 최적화하여 효율적인 자원 분배 방식을 제안한다.

다중객체의 객체별 중요성은 첫째로 산술적 특성, 즉, 객체의 움직임, 부호화되는 객체의 크기를 부호화 단위인 매크로블록 수로 표현한 것, 그리고 이전 프레임을 부호화한 결과 오차를 고려한다. 그러나 다중객체를 효과적으로 부호화하기 위해 객체별 중요도를 평가하여 중요도에 따른 목표비트를 설정하고 비트량을 분배하는 기법을 제안하였다. 제안된 MVO(multiple video object) 율제어는 현존하는 SVO(single video object) 기법의 확장이다. MVO 기법은 SVO 블록도에 네 부분을 첨가하여 구성되는데, 목표치 분배, 사전프레임스킵 제어, 동작모드 절제, 형상부호화 제어(AlphaTH 계산)이다. 첫 번째는 분리된 객체에 대한 별도의 율제어를 담당하고, 둘째는 영상정보가 비트량의 많은 부분을 차지할 때 버퍼를 안정화시키기 위한 것이다. 다음 두 성분은 공간 및 시간부호화 해상도간, 형상과 텍스처부호화 정확도간 트레이드오프를 제공하며 낮은 비트율에서 시스템 성능을 향상시키기 위한 것이다.

논문의 구성은 다음과 같다. 제2장에서 MPEG-4 MVO 율제어를 기준으로 다중객체 부호기에 관련된 율제어 기법을 설명한다. 제3장에서는 다중객체의 객체별 중요성을 크기, 움직임, 왜곡으로 표현하여 전체 목표비트를 객체별로 분배하는 기법을 제안한다. 제4장에서 다양한 테스트 조건하에 실험을 행하였다. 이전 복원 결과에 따라 목표치를 다중객체 전체에 대해 설정하고 적절히 분배한 결과를 보

이다. 끝으로 제5장에서는 논문의 주요 결과를 요약하고 향후 나아갈 방향을 제시한다.

II. 다중객체 부호기

그림 1에 다중객체 율제어 기법을 위한 블록도가 보인다. 크게 3단계 블록으로 나누어지는데, 사전부호화, 부호화, 사후부호화이다. 실제로 율제어 기법은 부호화 엔진에 어떤 변화를 가져오지는 않고 단지 사후부호화 단계에서 얻어진 데이터를 근거로 입력 QP를 부호화 엔진에 제공하는 역할을 한다. 여기서 다음의 정보가 분석된다. 즉, 현재 프레임에 적용된 QP, 이것으로부터 초래된 비트수, 전송되는 전체 비트수이다. QP와 텍스처 비트수를 이용하여 모델변수를 결정하고 전송되는 전체 비트수는 버퍼레벨을 갱신하기 위해 이용된다. 갱신된 버퍼량이 너무 많으면 사후프레임스킵(postframeskip) 제어기에 의해 적절한 수의 프레임이 건너뛰어진다.

사후부호화 단계에서 얻어진 정보를 사전부호화에서 이용한다. 우선, 이전 프레임에 사용된 비트수와 가용 비트에 기초하여 초기목표치를 산출한다. 이것은 버퍼충만도에 따라 조정되고 최종 QP가 계산된다. 최종 계산은 가장 최근의 사후부호화 단계에서 결정된 현재 모델변수에 근거한다.

이러한 과정은 단일객체 부호화나 마찬가지로이다. 다중객체의 경우 목표치를 예측하여 n 개의 객체에 적절히 분배해주어야 하며 이때 스킵될 프레임이 얼마인지를 고려한다(사전프레임스킵 제어). 다중객체는 일반적으로 임의의 형상을 갖는 객체의 집합이므로 단일객체와 달리 형상정보를 부호화할 필요가 있다. 이 경우 형상에 대한 율제어가 요구되며 형상과 텍스처간 비트율 분배도 주의해야 한다. 비트율이 낮으면 형상을 위한 비트가 상대적으로 크게 되며 반대의 경우에는 낮게 된다. 이와 같이 비트율의 분배는 다중객체 부호화에서 가장 중요한 요소로 대두된다.

다중의 객체는 의미적으로 상호 관련되어 있고 화질은 VO들의 종합된 화질로써 복잡한 함수가 된다. 객체들은 공통자원 즉 채널을 공유하므로 성능의 최적화는 전역 제어를 요구한다. 입력되는 n 개의 VOL은 각각의 특성과 중요도를 제시한다. 전체

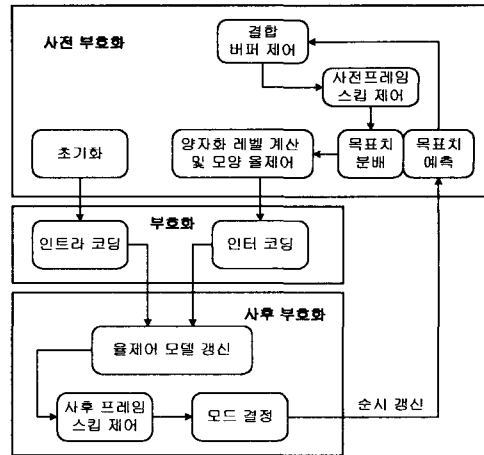


그림 1. 다중객체 부호화를 위한 블록도
Fig. 1. Block diagram for coding multiple video objects

시퀀스의 목표 비트율은 물리적인 채널용량에 따르며 주어진 비트율이 각 VOL에 배분된다. 이때 VOL의 시각적인 관심도는 배분에서 중요한 변수로 도입되어야 한다. VOL 부호기에서 생성된 비트량은 버퍼에 모아져서 전역적으로 제어된다.

1. 버퍼 초기화

버퍼에 관한 사항이 정의되고 부호화 파라미터가 초기화된다. 일반적으로 버퍼 크기는 주어진 비트율의 절반으로 한다. I 프레임을 부호화하기 위해 QP의 초기치는 입력변수로 주어진다. 첫 프레임을 부호화하는 데 사용된 비트수 T_f 를 제외하고 시퀀스의 나머지 프레임을 위해 가용한 비트수 b_r 을 결정한다. 즉,

$$b_r = t_s \cdot R_s - T_f \dots\dots\dots (1)$$

여기서 t_s 는 시퀀스의 주기를 초로 나타내고 R_s 는 시퀀스의 원하는 비트율이다. 또한 하나의 VOP가 부호화된 후 버퍼로부터 제거되는 비트수를 다음과 같이 정의한다.

$$R_{drain} = \frac{b_r}{n_r} \dots\dots\dots (2)$$

여기서 n_r 는 I 프레임 후 부호화되는 P 프레임

수를 의미한다. 후속 프레임을 위해 평균적으로 균등한 비트 공간을 마련하는 것이다. 이들 변수는 각각 MVO마다 주어진 비트율에 따라 다르게 정의된다.

2. 초기 목표 비트율 예측

이전 프레임의 i -번째 객체에 사용된 비트율 $R_{p,i}$ 와 비례하여 i -번째 객체를 위한 비트율을 배정하는 방식으로 목표치가 설정된다. 즉,

$$T_i = \max \left\{ \frac{R_s}{M \cdot F_s}, (1 - w_p) \frac{T_r}{M \cdot N_r} + (w_p) R_{p,i} \right\} \dots \dots \dots (3)$$

여기서 하중치 R_s 는 시퀀스 비트율, F_s 는 소스의 프레임율, T_r 는 가용 비트수, N_r 는 부호화될 잔여 P 프레임 수, M 은 VO 수를 각각 의미한다. 첫 항은 주어진 비트율을 VO 수로 나누고 프레임율로 나누어 VO당 평균비트가 계산된다. 둘째 항은 부호화되는 나머지 프레임당 평균비트와 이전 프레임에 사용된 비트수를 조합하는 것이다. w_p 값의 증가는 개별 목표치를 $R_{p,i}$ 에 더욱 비례하도록 만든다. 실험에서는 $w_p = 0.25$ 가 사용되었다.

3. 결합 버퍼 제어

초기 목표치가 결정되면 현재 버퍼레벨 B_c 와 버퍼크기 B_s 에 근거하여 조정된다. 전체비트 T_s 는 VO별 목표비트 T_i 의 합이다. 즉,

$$T_{joint} = T_s \cdot \frac{B_c + 2 \cdot (B_s - B_d)}{2 \cdot B_c + (B_s - B_c)} \dots \dots \dots (4)$$

이러한 조정은 매 프레임을 부호화한 후 버퍼점유도를 약 50%로 유지하도록 한다. 현재 프레임이 버퍼 크기에 맞게 부호화되었다면 전체 VO를 위한 버퍼는 목표비트의 절반이 된다. 100%가 넘으면 오버플로우가 되며 0% 이하이면 언더플로우되는 현상이 영상정보의 변화에 따라 발생할 수 있다. 이를 피하기 위해 추가로 최종 목표치를 다음과 같이 조정한다.

$$T = \begin{cases} (1 - \delta)B_s - B_c, & \text{if } B_c + T_{joint} > (1 - \delta)B_s \\ R_{drain} - B_c + \delta B_s, & \text{if } B_c - R_{drain} + T_{joint} < \delta B_s \\ T_{joint}, & \text{otherwise.} \end{cases} \dots \dots \dots (5)$$

버퍼상태가 최고레벨보다 크면 일정량을 감소시키며 반대로 너무 적으면 증가시키는 조정기법이다. 대표적으로 안전마진 δ 값은 0.1이나 낮은 비트율에서 영상 정보를 부호화하기 위해 0.2~0.25로 증가될 수 있다.

4. 목표치 분배

결합버퍼제어에서 결정된 비트를 VO마다 분배한다. 이 단계에서 목표 비트수 $T_i, i \in M$ 을 만족하도록 각각의 임의의 영상 VO에 결합 버퍼제어의 출력목표치가 분배된다. 영상의 크기, 움직임벡터의 크기, 그리고 분산량(왜곡)의 크기에 따라 객체간 분배가 이루어진다. 주어진 목표치에서 객체 i 의 목표치는 다음과 같다.

$$T_i = T \cdot (w_s SIZE_i + w_m MOT_i + w_v VAR_i) \dots \dots (6)$$

여기서 $SIZE_i, MOT_i, VAR_i$ 는 각각 객체 i 의 크기, 움직임, 및 MAD^2 이며 모든 객체의 전체 $SIZE, MOT, VAR$ 에 의해 정규화된 것이다. 정규화 이전의 i -번째 VO에 대한 변수는 각각 다음과 같이 정의된다.

$$\begin{aligned} MOT_i &= (abs(MV_x) + abs(MV_y)) / N_{MB}[i] \\ SIZE[i] &= N_{MB}[i] \\ MAD^2[i] &= (error_{sq}[i]) / N_{MB}[i] \end{aligned} \dots \dots \dots (7)$$

따라서 i -번째 객체의 움직임 크기 MOT_i 는 각 움직임벡터 성분의 절댓값이며 크기 $SIZE_i$ 는 객체에 포함된 매크로블록 또는 부분 매크로블록 수이다. 객체는 임의의 영상이기 때문에 경계면에서 구형 매크로블록과 부분적으로 형성된 블록이 존재한다. 하중치 $\{w_s, w_m, w_v\} \in [0, 1]$ 은 $w_s + w_m + w_v = 1$ 을 만족한다.

그러나 다음의 문제점을 갖는다. 영상의 크기는 배경 VO에서 가장 크게 되며 가장 많은 비트가 할당된다. 물론 움직임이 적다는 가정하에 전체 비트는 줄어들 수 있으나 영상 크기의 면에서 배경에 과다한 비트가 할당되는 문제점을 갖는다. 움직임에서도 만일 특정 객체가 계속 큰 움직임을 갖는 경우 계속 많은 비트가 할당되게 된다. 그러나 지속적으로 움직이는 물체의 부호화는 갑자기 움직임이

시작되거나 끝난 객체에 비해 시각적으로 덜 중요하다는 점을 간과하고 있다. 이 문제의 해결책을 다음 장에 논의한다.

III. 시각적 비트율 분배

형상의 크기를 전체 VOP 크기로 나누어 비율에 의해 부호화하면 결과적으로 형상의 크기가 크면 비례적으로 많은 비트가 할당되는 문제점을 낳는다. 이를 해결하려면 우선 형상의 크기를 시각의 공간 주파수 감도 특성으로 표현한다. 가장 큰 크기는 전체 영상을 포함하는 객체 즉 배경과 같은 경우에 해당하며 가장 낮은 공간주파수를 의미한다. 낮거나 높은 공간주파수에 대해 인간의 감도는 낮고 중간 주파수(약 5내지 10 cycles/degree)에서 감도가 높은 특성을 보인다[9]. 객체의 크기가 작으면 공간주파수는 높아지고 감도는 떨어지며 대체로 인물에 해당하는 중간 크기의 객체의 감도가 높게된다(그림 2). 따라서 형상에 대한 하중치는 크기에 따른 방사상 주파수를 식 (8)에 대입하여 적응적으로 얻어진다. 공간주파수는 관측거리에 따라 상대적으로 변하는 것으로서 거리가 가까우면 공간주파수는 낮아진다. 따라서 주어진 해상도의 영상에 대한 가장 낮은 주파수 기준을 정하여 가장 높은 주파수는 하나의 화소가 반 사이클을 차지하는 경우로 해석한다. 전체 영상을 위한 기준 주파수가 정해지면 그에 따라 관측거리를 선정하고 객체별 공간주파수가 계산된다.

$$H(f) = 2.6(0.0192 + 0.114f) \exp(-0.114f)^{1.1} \dots\dots\dots (8)$$

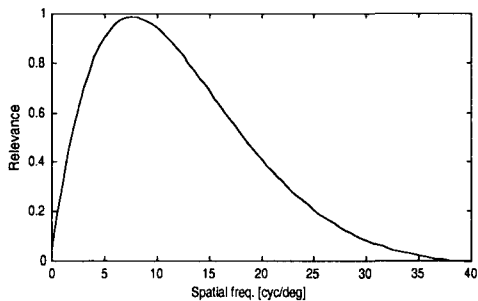


그림 2. 인간 시각시스템의 공간주파수 특성
Fig. 2. Spatial frequency characteristics in human visual system

움직임의 크기에 대한 중요성도 감도특성으로 설명된다. 공간주파수 대신에 시간주파수로 표현되며 마찬가지로 중간 정도의 움직임에 대해 높은 감도를 갖는다. 또 다른 특징은 움직임이 시작되었거나 멈춘 경우에 높은 감도를 갖는다는 것이다. 따라서 이 두 가지 특성을 이용하여 두 번째 하중치가 결정된다. 왜곡량은 일반적으로 절대 크기 자체로 이용한다. 상기와 같이 세 가지 변수에 대한 적응적 값이 결정되면 목표치의 전체 합을 일정히 하기 위한 기준 즉 하중치의 합이 1이 되게 하기 위해 비율을 재조정한다.

목표비트 분배의 마지막 단계로서 제안하는 것은 현재 프레임에 대한 특성뿐만 아니라 과거 프레임의 결과도 매우 중요한 것이다. 즉 형상, 움직임, 분산량은 과거 프레임에서도 존재하였으며 현재 프레임의 결과와 비교하여 차이값(DM, DS, DV)을 살펴보면 예를 들어 움직임이 시작되었는지 멈추었는지를 알 수 있다. 단 시작된 경우나 멈춘 경우 모두 우리는 대칭적으로 (같은 값으로) 가정한다. 차이값을 이용한 목표비트의 재분배는 다음과 같이 얻어진다.

$$T_i = T_i \frac{DM_i + DS_i + DV_i}{\frac{1}{num VOs} \sum_j (DM_j + DS_j + DV_j)} \dots\dots (9)$$

각 객체에 대한 전체 비트수가 결정되면 텍스처를 위한 가용 비트수는 움직임, 형상, 다른 부가정보를 제외하고 얻어진다. 양자화 변수는 앞에 언급한 바와 같이 각 객체에 대한 윌-양자화 모델을 이용하여 얻는다.

IV. 실험 및 결과

윌-왜곡 함수를 이차 윌제어 모델로 근사화하여 다중객체 부호화에 적용하였다. 실험에 사용된 시퀀스의 객체 수는 Akiyo(2), News(4), Coastguard(4)이다. 실험에 이용된 시퀀스와 객체 중에 "Coastguard" 영상에 대해 그림 3에 보인다. 모두 176 x 144의 QCIF, 30 frames/sec의 영상이며 총 길이는 300 프레임으로 10초간 분량이다. 이들은 MPEG-4에서 시험 영상으로 제공된 것이다. 객체 수는 버퍼레벨 제어에서 어려움을 주는 요소이며 객체의 의미에 따

라 주관적으로 결정된다. Akiyo는 뉴스룸에서 정적인 배경에서 앵커가 말하는 시퀀스이므로 2개의 VOL을 갖는다. News 영상은 배경, 춤추는 댄서의 삽입된 비디오, 두 명의 앵커, 그리고 텍스트로 구성된다. Coastguard 영상은 거의 정적인 바닷물 배경, 왼쪽에서 진입하는 배, 화면 중간에서 왼쪽으로 이동하는 조그만 보트, 그리고 상단에 해안 배경은 시간이 흐를수록 커지기 때문에 관심도가 증가되는 반면, 보트는 이동하여 사라지기 때문에 관심도가 감소되는 상태로 진행된다. VOL은 이미 사전에 자동 또는 수동의 방법으로써 분할된 것으로 가정한다.

다중객체의 객체별 중요성은 첫째로 산술적 특성, 즉, 객체의 움직임, 부호화되는 객체의 크기를 부호화 단위인 매크로블록 수로 표현한 것, 그리고

이전 프레임의 부호화한 결과 오차를 고려한다. 그러나 다중객체를 효과적으로 부호화하기 위해 객체별 중요도를 평가하여 중요도에 따른 목표비트를 설정하고 비트량을 분배하는 기법을 실험하였다. 실험에 사용된 시스템은 MPEG-4 VM8을 기초로 하여 객체별 중요도를 고려하였다. 움직임 크기는 각 움직임벡터 성분의 절대합이며 객체의 크기는 임의 형상 객체에 포함된 매크로블록 수로 표현된다. 경계면에서 매크로블록이 완전히 포함되지 않고 부분적으로 포함되는 것도 계산에 산입한다. 실험을 위한 변수를 표 1과 같이 설정하였다.

네 개의 객체를 가진 Coastguard 영상의 객체별 크기를 제시하는데, 다른 시퀀스와 달리 객체들이 복잡하게 변화한다. VO1은 바닷물 배경을 의미하지만 카메라 앵글이 팬닝하기 때문에 움직이는 물체로 인식한다. VO2는 왼쪽에서 조그맣게 존재하다가 시간이 흐를수록 가운데로 이동하여 끝까지 화면 중심에서 큰 물체로 존재한다. 역시 오른쪽으로 계속해서 움직이는 것으로 인식한다. 반면에 VO3는 화면 중간에 위치하다가 왼쪽으로 이동하는 물체이다. VO4 또한 화면 상단에 정지한 배경이지만 중간 부위까지 하강하는 움직임을 보이고 팬닝에 의한 오른쪽 방향의 움직임도 존재한다. 이처럼 복잡한 움직임을 보이며 정적인 물체는 존재하지 않는다. 그림 4에 이러한 흐름을 매크로블록 수로 표현한 크기의 함수로 보이는데 부호화된 프레임 28번까지 계속 복잡하게 변화한다. 크기가 커지면

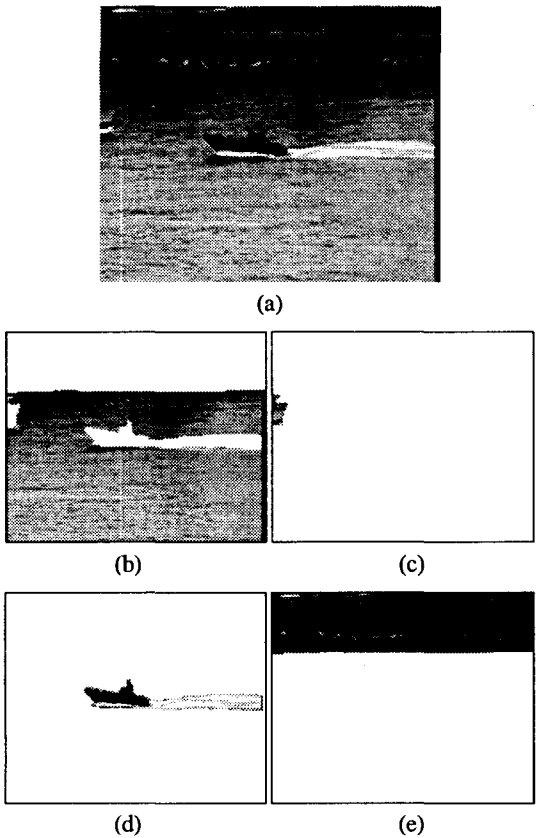


그림 3. 실험에 이용된 Coastguard 영상과 네 개의 객체
Fig. 3. Simulation sequence Coastguard and four video objects

표 1. 실험을 위한 변수 설정

Table 1. Parameter selection for simulation

변수	설정값
시퀀스	MPEG 표준 시퀀스
객체 수	4
전송률	128 kbps, 192 kbps
프레임율	10 fps(원영상은 30 fps)
프레임 수	200 프레임
해상도	QCIF(176 x 144)
QP 초기치	16
알파 초기치	64
버퍼 크기	64 kbits, 96 kbits
양자화	MPEG-2 형식
부호화 형식	IBBPBBP...
프레임 스킵	허용

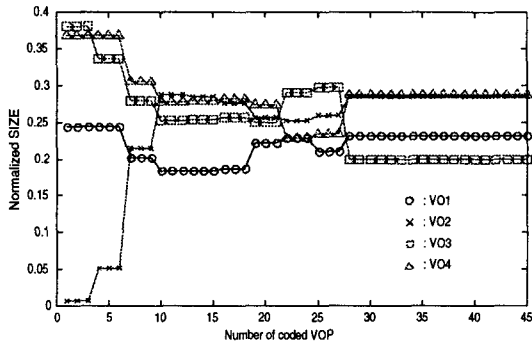


그림 4. Coastguard 영상의 객체별 크기 변화
Fig. 4. Object size variation of Coastguard sequence

보다 많은 비트를 요구할 수 있으나 앞에서 논한 바와 같이 변화량이 큰 객체에 보다 많은 가중치를 부여하였다.

그림 5에 제시한 움직임의 크기 또한 변화가 심하다. 바닷물에 해당하는 VO1의 움직임이 대체로 큰 움직임을 보이는 것에 주목한다. 중심부에 위치한 VO2, VO3보다 많은 비트를 할당받게 만드는 요인이다. 따라서 이 논문에서 제시하는 움직임의 차이 즉 변화량을 적용해야 하는 근거가 된다. VO4는 대체로 움직임이 적으나 중간에 위에서 하강하는 시점에서 상당한 비트량을 요구한다. VO1에 비해 전체적인 움직임은 적으나 내부의 복잡성으로 인해 많은 비트가 필요한데도 단순한 객체간 비율로는 이런 필요성에 쉽게 대처할 수 없다.

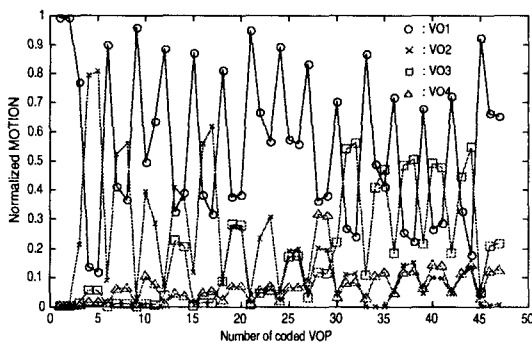


그림 5. 네 개의 객체를 가진 Coastguard 영상의 움직임 벡터 크기(128kbps)
Fig. 5. Motion vector magnitude for Coastguard sequence(128kbps)

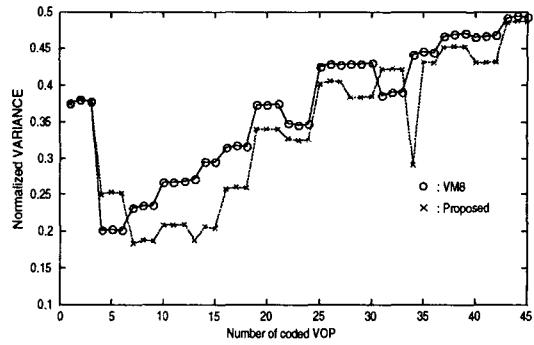


그림 6. 자승 MAD로 표현한 Coastguard 영상의 VO2 분산량
Fig. 6. Variance comparison of VO2 of Coastguard sequence represented by squared MAD

객체의 크기와 움직임의 크기와는 달리 부호화된 프레임의 왜곡량은 부호화 기법에 따라 변화한다. 이전 프레임의 결과를 가지고 다음 프레임을 부호

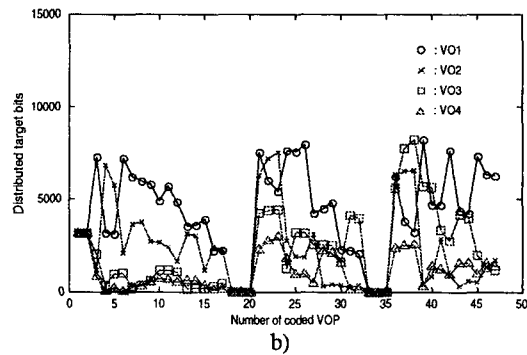
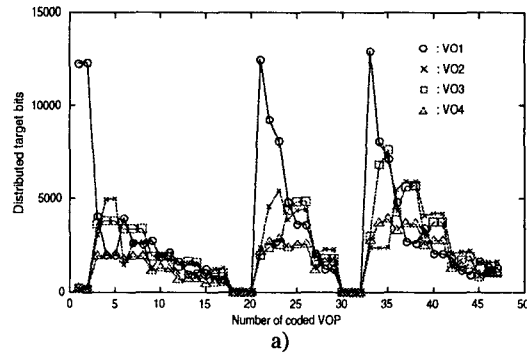


그림 7. 목표비트의 객체별 분배(128kbps의 Coastguard 영상): a) VM10, b) 제안된 기법
Fig. 7. Target bits distribution(Coastguard sequence at 128kbps): a) VM10, b) proposed scheme

화하는 쾌환이 이루어지기 때문이다. VO2를 중요한 객체로 인식하여 MPEG-4 VM10에서 채택한 기법과 제안한 기법의 결과를 그림 6에 제시한다. 정규화된 분산량으로 표현하여 최대 0.2의 감소량을 보인다.

그림 7에 각 VO에 할당되는 초기 목표비트를 제시한다. 단순한 비교를 이용하는 MPEG-4 VM의 결과는 VO1에 지나치게 많은 비트를 할당하는 경우를 볼 수 있다. VO1이 배경에 해당하지만 움직임이 존재하고 크기가 작지 않기 때문에 발생하는 현상이다. 그림 7 b)에서 제시하는 결과는 고르게 비트를 분배함을 보인다. 이처럼 객체의 중요성에 비추어 적절한 비트를 분배하는 데에는 크기, 움직임, 왜곡량이 프레임간 부호화 과정에서 절대량과 변화량을 이용해야 한다.

V. 결 론

다양한 종류의 비트율과 공간 및 시간해상도에서 다중 객체를 효과적으로 부호화하는 데에 필요한 비트량 분배기법을 제안하였다. SVO 기법에 첨가하여 MVO로 확장하는 시스템에 이용될 수 있다. 주요 첨가물은 객체간 목표비트 할당이다. 사전부호화 단에서 움직임 및 형상부호화한 후에 가용 비트를 분석하지만 버퍼상태와 시공간해상도가 버퍼 안정을 위해 조정된다. 또한 시간해상도에 근거하여 동작모드를 선택함으로써 율제어를 적용시킨다. 서로 다른 부호화 조건에 적용토록 하며 공간 및 시간해상도간, 형상과 텍스처 왜곡간 트레이드오프를 얻도록 비트율을 분배하는 기법을 제안하였다. 객체간 중요성을 분석하여 VOP로 나뉘어지고 객체기반 모델링에 의해 비트가 할당된다. 객체간 목표치 분배는 움직임벡터 크기, 객체의 크기, 이전 프레임의 왜곡 등을 반영하는데 단순한 비교에 의한 분배보다 변수의 변화량을 시각적으로 중시하여 목표치를 분배하였다. 제안한 기법을 MPEG-4 VM10 부호기에 사용된 기법과 비교하였다.

현재 사용된 율-양자화기 모델이 텍스처 부호화에 꽤 적절하지만 형상에 효과적인 R-D 모델 개발이 아직도 요구된다. 그러한 모델은 형상의 기하학적 상태에 의존할 것이다. 개별 R-D 모델외에 형상

및 텍스처 왜곡에 대한 감각적 왜곡의 이해가 형상 및 텍스처 결합 율제어를 위해 요구된다. 또한 서로 다른 객체가 다른 프레임율로 부호화되도록 하여 합성문제를 해결할 도구가 요구된다. 그러면 더 복잡한 버퍼링 기법을 필요로 하지만 가능성은 실로 크다.

참 고 문 헌

- [1] M. R. Pickering and J. F. Arnold, "A perceptually efficient VBR rate control algorithm," *IEEE Trans. Image Processing*, vol. 3, pp. 527-532, Sept. 1994.
- [2] P. Fleury, J. Reichel, and T. Ebrahimi, "Image quality prediction for bit rate allocation," in *Proc. IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, pp. 339-342, Oct. 1996.
- [3] X. Marichal, T. Delmot, C. DeVleeschouwer, V. Warscotte, and B. Macq, "Automatic detection of interest areas of an image or of a sequence of images," in *Proc. IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, pp. 371-374, Oct. 1996.
- [4] J. B. Lee and A. Eleftheriadis, "Spatio-temporal model-assisted compatible coding for low and very low bit rate video-telephony," in *Proc. IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, pp. 429-432, Oct. 1996.
- [5] L. Wang and A. Vincent, "Joint rate control for multi-program video coding," *IEEE Trans. Consumer Electron.*, vol. 42, pp. 300-305, Aug. 1996.
- [6] MPEG video group, "Text of ISO/IEC FDIS 14496-2," Doc. ISO/IEC JTC1/SC29/WG11 N2502, Atlantic City, NJ, Oct. 1998.
- [7] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 19-31, Feb. 1997.
- [8] A. Vetro and H. Sun, "Joint rate control for coding multiple video objects," in *Proc. IEEE Workshop Multimedia Signal Processing*, Princeton,

NJ, pp. 181-186, June 1997.

- [9] J. L. Mannon and D. J. Sakrison, "The effects of a fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. 20, pp. 525-536, July. 1974.



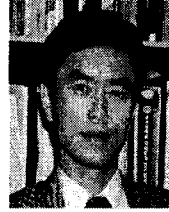
지 석 상(Seok-Sang Chee)

2000. 8. 전북대학교 전자공학과
공학박사

1988. 2. 전북대학교 전자공학과
공학석사

1983. 2. 전북대학교 전자공학과
공학사

1988. - 1995. 한국전자통신연구원 연구원



황 재 정(Jae-Jeong Hwang)

군산대학교 전자정보공학부 교수

전북대학교 전자공학과 공학박사

텍사스주립대학교 객원교수

한독기술협력사업 파독교수

군산대학교 공학연구소장



이 문 호(Moon-Ho Lee)

일본 동경대 전자과 공학박사

전북대 정보통신공학과 교수 및

정보통신연구소 소장

한국통신학회 상임이사, 한국공학

원 한림원 회원

정보통신부 정책심의위원