

동영상 부호화를 위한 의미 기반 Rate Control 기법

정희원 이봉호*, 전경재*, 콧노윤**, 강태하**, 황병원**

A Semantic-based Rate Control Method for Motion Video Coding

Bong-Ho Lee*, Kyung-Jae Jeoun*, No-Yoon Kwak**, Tae-Ha Kang**, Byong-Won Hwang**
Regular Members

요 약

본 논문에서는 비트율이 매우 한정된 응용에 적용되는 초저속 동영상 부호화 표준인 H.263plus에 기반한 의미 기반 율-제어 기법을 제안한다. 기존의 율-제어 기법들은 의미기반이 아닌 단순히 프레임에 대해 매크로 블록 단위로 최적의 양자화 파라미터를 설정해 정해진 비트율을 조정하나, 본 논문에서는 부호화 효율을 높이기 위해 의미 영역 인식 및 우선권 부여를 통해 영역별로 최적의 양자화 파라미터를 설정함으로써 제약된 비트 내에서 주관적인 화질의 향상을 보이는 기법을 제안한다. 제안하고자하는 기법의 핵심은 부호화 시 배경에 해당하는 영역에 배정되는 비트량을 할당해 전경(얼굴) 영역에 비트 자원을 더 많이 배정함으로써 주관적인 화질을 보호, 개선하기 위한 것이다.

ABSTRACT

This paper presents the semantic based rate-control method which is based on very low bit rate video coding standards H.263plus, applied on very low bit rate applications. Previous rate control methods control the generated bit rates by setting the optimum quantization parameters per macro block unit on frame. But, in this parer, we added the pre-processing algorithm, semantic region recognition and assignment of priority algorithm, to obtain the subjective quality enhancement. This work aims to improve the subjective quality of skin color region or face by using unimportant background region's bit resources.

I. 서론

앞으로 도래할 멀티미디어 시대의 핵심이 되는 통신, 방송, 컴퓨터 기술은 디지털 기술의 발달로 인해 급속히 변화하고 있다. 통신부분에 있어서는 현재, 보다 나은 수준의 통신 품질을 요구하는 사용자들의 요구에 부합하는 영상 및 멀티미디어 통신 기술이 연구되고 있다. 화상 회의나 화상 전화를 위한 동영상 압축 기법으로는 H.26x 계열로 표준화가

진행되어 왔고, H.261은 64kbps~2Mbps 범위의 비트율을 갖는 응용 예인(N-ISDN)에 적합한 화상 전화 및 화상 회의용으로 표준화가 완성이 되었다^[1]. 또한 저속 동영상 부호화(row bitrate video coding) 표준이라 할 수 있는 64kbps 근처의 대역폭이 요구되는 화상 통신을 위한 압축 표준으로 H.263이 1995년에 완료가 되었고 초저속 동영상 부호화 표준으로 대표되는 64kbps이하의 대역폭을 갖는 이동통신망(GSM)이나 PSTN망을 이용한 화상전화(video telephony)응용을 위한 동영상 압축 표준으로

* 김포대학 전자정보계열(jslee@kimpo.ac.kr)

** 한국원자력연구소 로봇기술개발실(yblee2@kaeri.re.kr)

*** 광운대학교 전자공학과(eskim@daisy.kwangwoon.ac.kr)

논문번호 : 99502-1224, 접수일자 : 1999년 12월 24일

H.263plus의 표준화가 진행이 되고 있는 상태이다.^[2] 현재 H.263plus는 성능 개선을 위해 H.261과 H.263 기반에 여러 개선된 움직임 추정 모드, BP 프레임 모드 및 향상된 율-제어 기법 등을 적용해 성능 향상을 보이고 있다.^[2]

영상 통신의 궁극 목적은 필요한 정보의 전달에 있다고 할 수 있다. 특히, 전송 대역폭이 매우 제한적인 응용의 경우에는 주어진 자원 내에서 양질의 화질을 유지하기 위해서는 강한 율-제어(rate control) 기법이 요구된다. 본 논문의 취지는 화상전화나 화상회의 응용 등에서 배경보다 중요한 의미를 갖는 전경 즉, 얼굴에 대해서 배경이나 나머지 영역에 배정되는 비트 자원을 할애해 이들 영역에 우선적으로 배정함으로써 낮은 비트 자원을 갖는 응용에 있어서도 주관적인 화질을 유지, 개선하기 위함이다. 기존의 율 제어 기법들은 의미 기반이 아닌 단순히 프레임에 대해 매크로 블록 단위로 최적의 양자화 파라미터를 설정해 정해진 비트율을 조정하나, 본 논문에서는 부호화 효율을 높이기 위해 의미영역 인식 및 우선권 부여를 통해 영역별로 최적의 양자화 파라미터를 설정함으로써 제약된 비트 내에서 얼굴 영역의 주관적인 화질의 향상을 보이게 하겠다. 본 논문의 구성은 다음과 같다. 먼저, II장을 통해 기본적인 비트율-제어 알고리즘을 설명하고 III장에서는 제안하고자 하는 인식 및 비트율-제어 기법을 설명하고, IV장에서는 실험결과 및 고찰을 다루며, V장에서 결론을 기술하기로 한다.

II. 율-제어(Rate Control) 기법

대부분의 통신 채널은 고정된 스트림을 입력으로 받아들이며, 가상 채널을 설정하는 ATM망에서도 스트림을 처리하는데 있어서 다소 제한을 가하고 있다.^[3] 이러한 특성으로 인해 고정 비트가 아닌 가변 스트림을 생성하는데 부호화 시 초래되는 성능 저하를 극복하기 위하여 율-제어 기법을 적용한다. 이 기법은 크게 실시간(on-line)과 비-실시간(off-line)으로 나눌 수 있는데 실시간 응용 여부에 따라 결정이 되며 MPEG1은 비-실시간, MPEG2, H.26x 계열 및 MPEG-4는 실시간에 해당되며 목표 비트율이 적은 응용일 수록 매우 효율적인 기법이 요구된다. 또한 MPEG과 H.263plus에 적용이 되는 기법에는 다소 차이가 있는데 이는 GOP의 길이와 구조가 다르기 때문이다. 이 장에서는 율제어 기법의 두 가지 방식인 고정 비트율 및 가변 비트율에 대

해서 알아보고, 기본적인 비트 배정 문제와 H.263plus의 TMN8 비트율 제어 방법을 간단히 소개하고자 한다.

2.1 고정 비트율(CBR)과 가변 비트율(VBR)

율 제어 영역은 크게 고정 비트율(CBR : constant bit rate)과 가변 비트율(VBR : variable bit rate)로 나누어지는데 CBR은 채널이 시간에 관계없이 항상 일정한 망에서 적용되는 비트율을 말하며 스트림을 전송하기 위해서는 정해진 시간당 즉, 해당 프레임들에서 발생하는 비트 스트림을 일정하게 유지시켜야 한다. 그러므로, CBR인 경우에는 주된 목표가 비트율이 되므로 화질을 희생해서 정해진 비트율을 조절해 주어야 한다. 다음으로, 가변 비트율(VBR)은 실시간적으로 채널이 변하는 망에서 제공 가능한 비트율을 말하며 이 경우에는 타겟을 화질에 맞추어 비트율을 조절하는 것이다.^[8] 현재는 VBR에 대한 연구가 많이 이루어지고 있고 H.263이나 MPEG-4 시스템에서는 이에 맞는 알고리즘이 개발된 상태이다.

2.2 비트 배정 문제(Bit Allocation Problem)

비트를 최적화 시키는 문제는 부호화 시 발생하는 에러를 최소화하고 정해진 비트에 가장 근접한 해를 찾는 것이다. 기본 식은 다음과 같다.

$$\sum_i R_i \leq Rc \quad \text{하에서} \quad (1)$$

$\text{Min } D_i$ 를 만족하는 해를 찾는 과정으로 이루어진다.

여기서, Rc : 정해진 타겟 비트량,

R_i : 양자화 후 발생 비트량

D_i : 양자화 후 발생 에러량

그림 1은 일반 부호화기의 부호화 제어(coding control) 과정을 나타낸 블록도로서 버퍼단으로부터 반환되는 버퍼정보 및 제어에 필요한 파라미터를 가지고 제어단에서 제어량을 결정하고 양자화단에서 양자화를 실행한다. 양자화식은 식 2와 같다.

$$y(i, j) = r \frac{8x(i, j)}{QW(i, j)} \quad (2)$$

여기서, $y(i, j)$ 는 양자화된 최종 계수값, $x(i, j)$ 는 8x8 DCT 계수, r 은 라운딩을 의미하며 Q 는 (1~31)까지의 이산 양자화 파라미터를 나타내며, 마지막으로 $W(i, j)$ 는 8x8로 고정된 양자화 매트릭스 계수를 나타낸다.^[4]

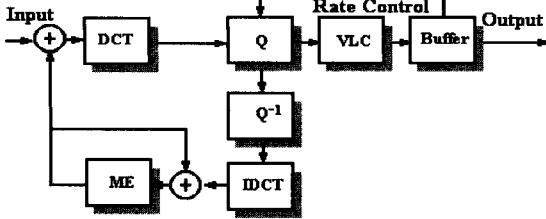


그림 1. 일반 부호화기의 부호화 제어블록도

2.3 H.263plus의 울-제어 기법

TMN8 에서 제공하는 울 제어 기법은 H.263plus 에 적용하기 위한 것으로서 부호화 효율을 향상하기 위해 IBP 구조에서 파생된 진보된 BP 프레임 구조(advanced BP frame structure)를 사용한다.^[2] 그러므로 프레임 계층 비트 배정 및 프레임내의 매크로 블록별 비트 배정이 기존의 MPEG-1, 2 시스템과는 다르게 적용된다.

프레임 계층 울 제어는 초당 정해진 프레임에서 버퍼의 상태 및 설계된 제어 시스템에 따라 해당 프레임을 스킵하기 위한 것으로서 실시간 적용을 위해 간단하게 구현이 이루어진다. 알고리즘 구현에 사용되는 제어 파라미터는 다음과 같다.

- W : 부호화 버퍼 자원
- B : 전 프레임에 의해 소모된 비트량
- R : 초당 정해진 타겟 비트율
- G : 원 비디오 신호의 프레임율
- F : 초당 프레임별 타겟 프레임율 (G/F는 정수)
- M : 프레임을 스킵하기 위한 임계값 ($M=R/F$)
- A : 타겟 버퍼 지연 시간은 $A \cdot M$ 으로

$$A=0.1.\text{default (sec)}$$

부호화 버퍼 $W=\max(W+B-R/F, 0)$ 로 skip = 1로 설정한다. 다음으로, 스킵 조건은 아래 코드에 따라 결정이 되며,

```
while (W > M) {
    W = max (W-R/F, 0);
    skip ++; }
```

입력되는 다음 비디오 시퀀스들 중에서 “skip·G/F-1”에 해당하는 프레임이 스킵된다. 프레임당 결정되는 타겟 비트율은 식 3을 통해 결정이 되는데, 결정된 이 비트율을 가지고 매크로 블록별로 최적의 QP (양자화 파라미터)를 설정해 울을 조절하게 된다^[5].

$$B = \frac{R}{F} - \Delta, \Delta = \begin{cases} \frac{W}{F}, & W > A \cdot M \\ W - A \cdot M, & \text{Otherwise} \end{cases} \quad (3)$$

다음 장에서는 제안 기법인 각 프레임 별로 배정된 타겟 비트율에 대해서 분할된 영역별로 우선권을 부여한 후 비트율을 조절하는 의미 기반 울 제어(semantic based rate control) 기법을 소개한다.

III. 의미 기반 울-제어 기법

3.1 개요

화상 회의나 화상 전화 같은 영상은 크게 사람에 해당하는 전경 (foreground image)과 배경(background image)으로 구성되는데 전경은 배경에 비해 의미가 크다고 볼 수 있다. 또한 전경을 얼굴(face)과 몸통(body) 부분으로 세분해보면 얼굴 부분이 의미를 더 가진다고 볼 수 있다. 이러한 점에 착안해서 움직임을 갖는 영역들을 분할하고 이 분할된 정보를 이용해 얼굴 영역과 손 영역을 인식해 제한된 채널 환경에서도 얼굴의 화질을 유지할 수 있다. 실제로 위에서 제시한 영역 분할 결과를 살펴보면 구성된 영역들은 얼굴 영역과 손을 중심으로 한 신체 영역으로 분할된다.

기존의 압축 표준안들의 비트 울 제어방법은 의미를 갖는 영역에 대해서 차별적으로 적용을 하지 않았다. 각기 요구되는 프레임구조 (IBP) 및 모델링 여부에 따라 다르게 적용이 되는데 각 프레임 내부는 매크로 블록 단위로 이루어지나 화질은 프레임 전체에 대해서 거의 동일하게 유지가 되어 배정할 비트 자원이 매우 낮을 경우에는 화질 저하가 심하게 나타난다. 이를 극복하기 위하여, 분할된 영역 중 의미가 큰 얼굴과 손을 중심으로 한 영역들에 우선권을 부여해 우선권이 높은 영역에 비트 자원을 많이 배정하고 우선권이 낮은 영역에 대해서는 비트 자원을 적게 배정함으로써 우선권이 높은 영역 즉, 화상전화나 화상회의에서 화자의 얼굴에 해당하는 영역의 화질을 주관적으로 보호 유지하며, 나아가서는 사용자에게 의해 지정되는 우선권 영역 (ROI : region of interface)에 대해서도 지원 가능한 의미 기반 비트 울 제안 기법을 제시한다.

3.2 의미 영역 인식 및 울-제어를 위한 전처리

3.2.1 영역 분할 (Segmentation)

영역 분할은 실제로 움직임을 갖는 블록들을 영역 별로 구성하기 위한 것으로서 병합(merging) 알고리즘을 적용해서 분할한다. 영역 병합 기법은 영역 성장 기법 중에서 가장 기본이 되는 것으로 각 화소를 단일 영역으로 고려해 $H(R) = TRUE$ 의 유사도 조

건을 만족하는 영역에 대해서 병합을 행한다. 일반적인 영역 병합은 아래 과정과 같이 이루어진다.

1. 유사도 조건 ($H(R_i) = TRUE$)에 만족하는 많은 작은 영역들로 이미지를 분할하기 위해 starting method를 정의한다.
2. 인접한 두 영역을 병합하기 위해 유사도 조건 (homogeneity criterion)을 정의한다.
3. 유사도 조건을 만족하는 인접 영역을 병합한다.
4. 마지막 영역까지 유사도 조건을 적용해 성장시키며 병합한다.

이는 가장 일반적인 영역 병합 기법으로 starting method나 merging criterion의 정의 여부에 따라 다양한 결과를 나타내며, 가장 간단한 starting method로는 2x2, 4x4, 8x8 화소로 구성된 블록을 설정한 후 통계적 파라미터(평균이나 분산)등을 이용해 유사도를 측정해서 병합 여부를 결정한다^[10]. 영역이 이웃하는 어느 영역과도 병합하지 못하면 그 영역의 병합은 중지하고 다른 영역을 병합해가며 모든 영역에 대해서 처리한다. 본 논문에서 적용한 영역 병합 알고리즘은 다음과 같다.

1. 두 프레임의 차 영상(difference image)을 구한다.

$$DiffPixel(x, y) = |I_k(x, y) - I_{k-1}(x, y)|$$

where, $I_k(x, y)$, current image

$$I_{k-1}(x, y), \text{ previous image} \quad (4)$$

2. 차 영상에 대해 8x8 블록으로 나눈 후, 차 화소 값이 15(15~20)보다 작은 화소들은 제거를 하고 또한 차 화소수가 1이하인 블록들은 제거한다.

3. 시작 방법(starting method)으로는 차 영상의 상단 좌측에서부터 시작하며 8x8 블록별로 8방향을 고려해서 병합을 수행한다. 그림 2에 예시된 병합 절차도에 따라 병합한다.

$$if(M(B_i) == TRUE)$$

$$RegionMerging();$$

여기서, $M(B_i)$ 는 블록 i 의 병합 조건

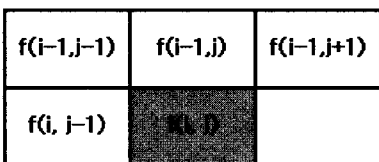


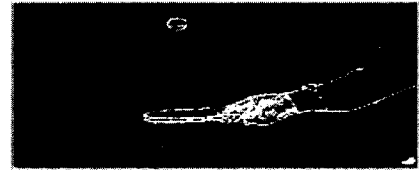
그림 2. 8방향 영역 병합 절차도

4. 병합시 차 영상에 나타나는 고립된 작은 블록들은 잡음에 의한 블록에 가깝고 그다지 영상에서 의미를 갖지 않으므로 제거한다. CIF 이미지인 경우에는 구성된 영역의 블록 수가 3보다 작을 경우에는 제거를 하고 QCIF 이미지인 경우에는 블록수가 1인 경우에 제거하였다. 이 과정은 실제 시뮬레이션 과정을 통해 검증된 바로 CIF 이미지인 경우에 거의 잡음에 의한 블록으로 생긴 영역이었으며 제거를 해도 거의 무방하였고 QCIF 이미지인 경우에는 이미지 자체가 down sampling 되었으므로 한 블록만으로 구성된 것만을 제거하였다.

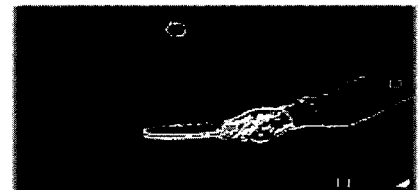
5. 마지막으로 구성된 영역들에 관한 레이블 정보를 추출한다.



(a) 원 영상



(b) 차분 영상



(c) 차분화소를 제거한 영상



(d) 소 영역을 제거한 후 분할된 영상

그림 3. 영역 분할된 이미지 (table tennis)

그림 3은 영역 분할을 보여주는 예로서 그림 3(a)는 원 Table Tennis 이미지를 나타내고 그림 3(b)는 차 이미지를 나타내며 그림 3(c)는 차 이미지에서 차 화소가 15보다 작고 고립된 화소들을 제거한 후의 이미지를 나타내며, 그림 3(d)는 병합 후 구성

step 1: 살색 화소 인식 : 그림 6에서의 타원 영역은 색 신호가 존재하는 영역으로 살색화소 결정 여부는 각 화소가 아래 식을 만족하면 살색 화소로 간주한다.

step 2: 살색 블록 결정 : 살색 블록 결정 여부는 블록 내부의 살색 화소수가 미리 정한 임계치 ((blocksize x blocksize x 3)/4) 이상일 경우에 살색 블록으로 결정한다.

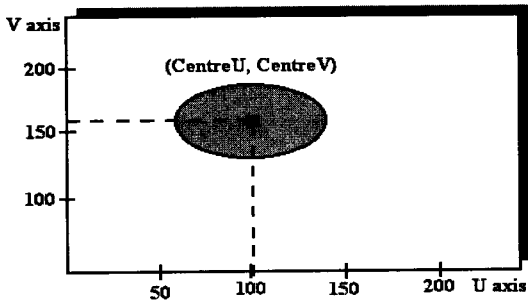


그림 6. UV 살색 좌표계

step 3: 살색 영역 결정 : 마지막으로, 살색 영역은 각 영역이 살색 블록을 하나 이상 가지고 있으면 살색 영역으로 결정한다.

3.3.1.2. 얼굴 영역의 분석 : 얼굴 영역은 일반적으로 화면 중앙에 위치하므로 가로축으로 창을 설정해 3.3.1.1 단계에서 찾은 살색 영역 중 창 내부에 속한 분할 영역(segmented region)들을 결정한다.

3.3.1.3. 세로축 창 설정을 통한 얼굴영역 결정 : 위 단계들을 통해서 판정된 살색 영역들이 얼굴이 아닌 손 같은 다른 영역일 수 있으므로 다시 세로축으로 창을 설정해 3.3.1.2에서 결정된 살색 영역 중 세로 축 창 내부에 속한 얼굴 영역을 결정한다. 그림 7은 창 설정 예를 도시하고 있다.

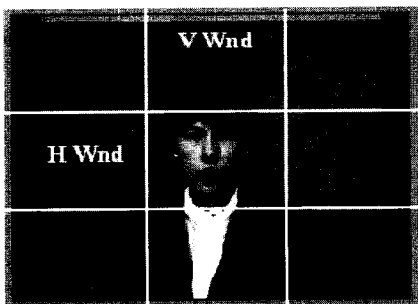


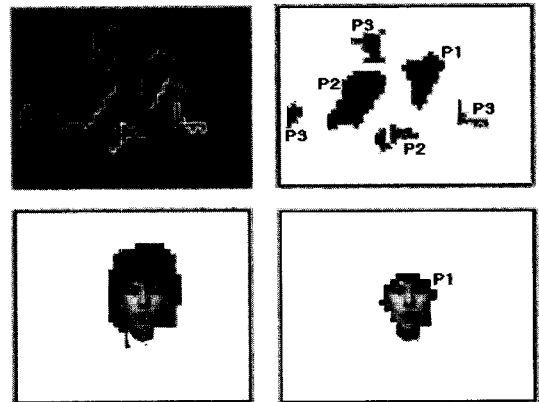
그림 7. 가로축 및 세로축 창 설정

3.3.1.4. 우선권 부여 : 이렇게 결정된 얼굴 영역에 대해서는 최 우선권을 부여하고 얼굴이 아닌 손 같은 살색 영역에 대해서는 중간 우선권을 부여하며 배경과 같은 의미가 덜한 영역에 대해서는 최저 우선권을 부여한다. 그림 8은 우선권이 부여된 분할된 영상을 나타낸다. 우선권을 부여하는 방법은 다음과 같다.

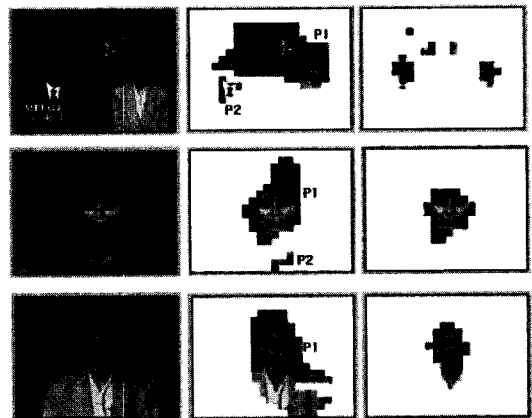
P1 [Face Region : High_Priority].

P2 [Non Face Skin Region : Middle_Priority]

P3 [Non Skin Region : Low_Priority]



(상) salesman (하) claire 이미지



(상) News (중) Miss america (하) Akiyo

그림 8. 인식 및 우선권이 부여된 이미지

3.3.2 의미기반 비트 율 제어 기법

율-제어 부호화기의 성능은 제어 시스템(control system)을 어떻게 모델링하느냐에 달려 있는데 크게 Buffer 모델링, 프레임 구조(IBP), 율-왜곡 함수 설정(rate-distortion function) 및 제어 함수(control

function)의 설정 여부에 따라 성능이 좌우된다.^[3] 실시간 처리가 요구되는 응용에서의 율을 조절하는 알고리즘은 계산상의 지연을 방지하기 위해 복잡도를 단순화 해야하고 최적의 화질을 유지하도록 설계가 되어야 한다.

3.3.2.1 움직임 보상 부호화

여러 보상시 Inter/Intra, MC/noMC 모드의 결정은 H.261에서 권고하는 방법^[1]을 이용하여 보상 부호화를 수행하고 발생된 총 비트량, 우선권이 정해진 영역별로 발생된 각 비트량과 우선권이 1인 영역에서 발생하는 살색블록의 발생 비트량 등을 구하고 이들을 율 제어 파라미터로 결정한다.

3.3.2.2 율 제어

지금까지의 과정에서 결정된 정보와 비트량을 이용해 최적의 비트를 배정한다. 비트 배정 절차는 낮은 우선 순위를 갖는 영역부터 처리한다. 그림 9는 율 제어를 나타내는 블록도이다. 이 과정을 대략적으로 살펴보면, 먼저 QP=0 DCT 단은 움직임 에러 보상을 위한 것으로 이 단에서 실제 발생된 비트량이 계산되며, R>T 이면 P3 영역을 위한 타겟 비트를 계산한 후 결정된 이 자원을 가지고 P3 영역에 대한 비트를 배정하게 된다. 다음으로 P3 영역에서 사용된 비트 자원을 가지고 P2 영역을 위한 타겟 비트를 계산한 후 다시 P2 영역의 비트 배정을 수행한다.

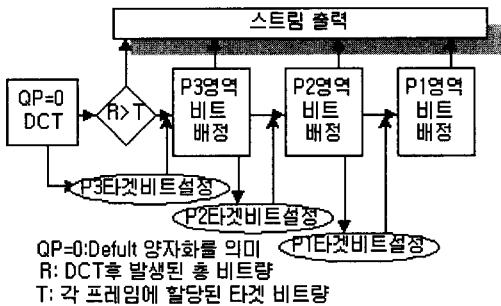


그림 9. 제한한 율 제어 블록도

마지막으로 P1 영역에 대한 비트 배정은 P3와 P2 영역에서 배정된 비트 자원을 뺀 나머지 자원을 가지고 수행하는데 이 단에서는 P1 영역 중 살색을 가지고 있는 영역을 우선적으로 보호하기 위해 비살색 영역에 대한 타겟 비트를 먼저 계산해 비트 배정을 하고 최종적으로 남은 비트 자원을 가지고 살색 블록들에 대한 비트 배정을 수행한다. 세부 진행 과정은 다음과 같다.

1. P3 영역 비트 배정 : 이 영역은 우선권이 가장 낮은 영역으로서 먼저 비트 배정을 수행한다. 과정을 살펴보면, 먼저 타겟 비트(Budgetbits)와 P프레임에서 발생된 총 비트량(Totalrealbits)을 비교하여 Totalrealbits이 크면 P1과 P2 영역을 우선적으로 보호하기 위해 P3 영역의 타겟 비트(P3Targetbits)를 구해 최적의 QP를 찾아 부호화한다. 이때 P3 영역을 위한 타겟 비트가 QP=31에 의해 발생된 비트량보다 적으면 최소 화질을 보장하기 위해 QP=31에 의해 발생된 스트림을 전송한다. P3 영역의 타겟 비트는 아래 식에 의해 계산이 된다.

$$P3Targetbits = Budgetbits - (P1realbits + P2realbits) \quad (6)$$

여기서, Pxrealbits : 우선권이 x인 영역에서 발생된 비트량

2. P2 영역 비트 배정 : P2 영역은 살색영역이면서 얼굴이 아닌 영역으로 P1 영역보다 우선권이 낮으며 이 영역을 위한 타겟 비트(P2Targetbits)계산은 Budgetbits에서 P3 영역에서 배정된 비트 자원을 뺀 나머지 자원 중 P1 영역에서 발생된 비트량을 뺀 나머지 자원을 가지고 비트 배정을 한다. 마찬가지로 P2 영역을 위한 비트 자원이 없을 경우에는 QP=31로 최저 화질을 보장한다.

$$P2Targetbits = (Budgetbits - P3codedbits) - P1realbits \quad (7)$$

여기서, P3codedbits: P3 영역 배정에 사용된 비트량

3. P1 영역 비트 배정 : P1 영역은 우선권이 가장 높은 살색 영역 중 얼굴에 해당되는 영역으로 P2와 P3 영역에 배정되고 남은 자원을 가지고 비트를 배정하는데 이 단에서는 P2, P3 영역과 달리 P1 영역 중 살색과 비살색 영역을 다시 구분해 비트 배정을 수행한다. 그 과정을 살펴보면 먼저 P1 영역 중 살색 블록들을 우선적으로 보호하기 위해 남은 자원(P1Targetbits) 중 살색 블록에 의해 발생하는 자원을 뺀 나머지 자원(P1nsrTargetbits)으로 비살색 블록들에 대해서 비트 배정을 한다. 이 때에도 배정할 자원이 없을 경우에는 QP=31로 최저 화질을 보장한다. 마지막으로 P1의 살색 블록들은 최종적으로 남은 자원 (P1srTargetbits)을 가지고 비트를 배정한다.

$$P1Targetbits = (Budgetbits - P2codedbits - P3codedbits)$$

$$P1nsrTargetbits = P1targetbits - P1srbits$$

$$P1srTargetbits = P1Targetbits - P1nsrcodedbits \quad (8)$$

여기서, P2codedbits: P2영역에 배정된 비트량
 P1srbits: P1 영역중 살색 블록들에 의해 발생된 비트량

P1nsrcodedbits: 비살색 블록에 배정된 비트량

4. **울 제어된 총 비트량** : 배정된 총 비트량 (Totalcodedbits)은 각 영역에 배정된 비트량을 합한 양으로 (Totalcodedbits = P1codedbits+ P2codedbits + P3codedbits)의 식으로 계산이 된다. 이때 배정된 총 비트량이 해당 프레임에 배정된 타겟 비트보다 클 경우에는 초과 비트량은 다음 프레임의 울 제어 파라미터로 정해지고 다음 프레임의 비트 울 제어 시에 영향을 미치게 된다.

$$\begin{aligned} \text{Totalcodedbits} &= \text{P1codedbits} + \text{P2codedbits} \\ &\quad + \text{P3codedbits} \\ \text{Overbits} &= \text{Totalcodedbits} - \text{Budgebits} \end{aligned} \quad (9)$$

최적의 비트 배정을 위한 QP 탐색 과정은 많은 처리 시간을 요하게 되는데 이를 극복하기 위해 그림 10과 같은 이진 QP 트리를 통해 빠르게 탐색을 수행한다.^[8] 한 영역에 대해서 1-31까지 32번을 탐색해야 하나 아래와 같은 이진탐색 방법을 이용해 탐색시 5-6번이면 탐색이 가능해 고속 수행이 가능하며 실시간 적용에도 유리하다.

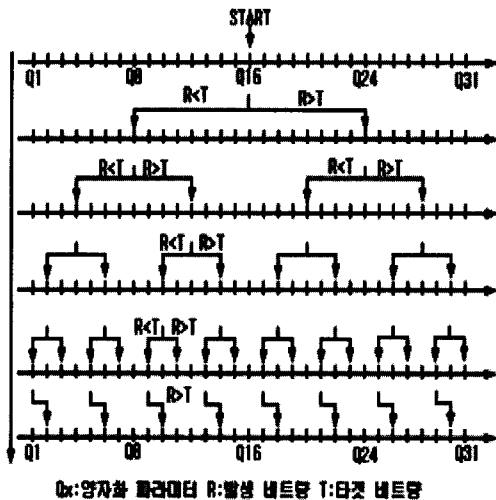


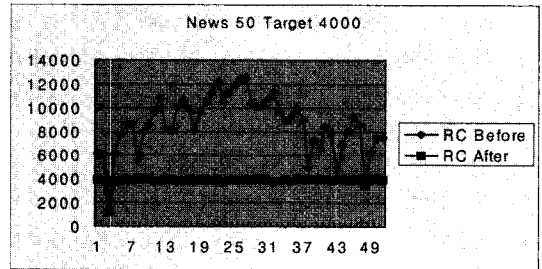
그림 10. 이진 QP 탐색 트리

IV. 시뮬레이션 결과 및 고찰

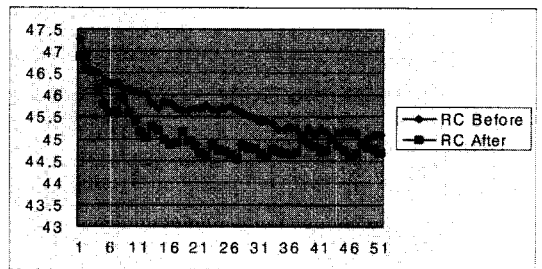
제안한 기법의 성능 평가를 위해 CIF Claire (타겟비트:10000)과 QCIF 크기인 MissAmerica (4000), News(4000), Akiyo(4000), Suzie(4000) 각

50 Frame에 대해서 타겟 비트를 설정하고 수행하였으며 그 결과는 그림 11~ 그림 14과 같다. 시뮬레이션 수행 정도는 각 프레임의 타겟 비트량을 가상으로 동일한 양으로 설정하고 P 프레임의 휘도 성분만 대해서만 모의 실험을 수행하였다. 그림 14에 예시된 차트 결과를 보면, 50 프레임의 Claire 영상은 10000으로 타겟 설정 시 타겟을 넘어서는 프레임에 대해서 타겟에 평균 9683.11으로 상향 근접한 비트 울을 얻을 수 있었고, 그림 11의 News는 3930으로 타겟 울에 상향 근접한 결과를 얻을 수 있었다.

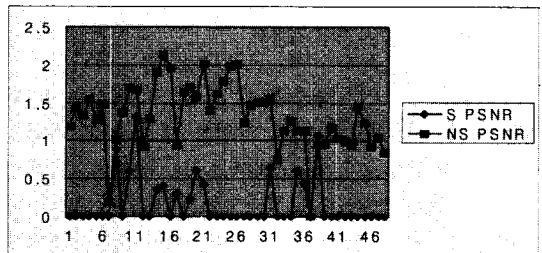
복원된 화질을 살펴보면, 표 1.에서 볼 수 있듯이 울 제어 후의 PSNR은 제어전보다 평균 0.27-0.70 dB 정도 낮게 나타났지만, 제어후의 PSNR 비교 결



(a) 울 제어 결과 (타겟 비트 4000, 50 Frame)



(b) 울 제어 전 후 PSNR



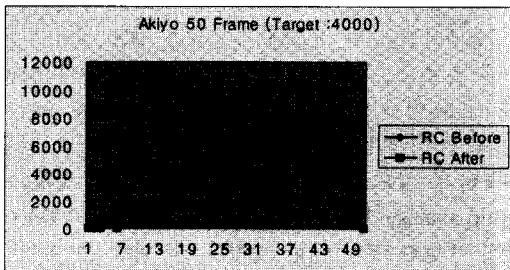
(c) 살색 영역과 비 살색 영역의 울 제어 전 후 PSNR 비교

그림 11. News 이미지

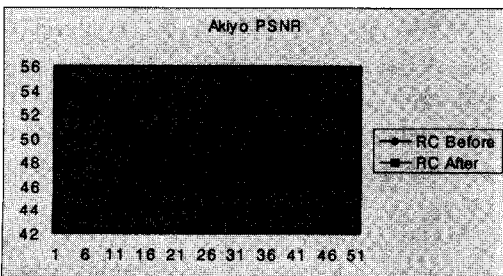
표 1. 울 제어 성능 비교표

시험 영상 (평균 PSNR)	울제어 전 PSNR	울제어 후 PSNR	살색 영역 PSNR 변화	비살색 영역 PSNR 변화	비고
MissAmerica	45.37	45.10	0.0149	0.2486	QCIF
News	43.9	43.25	0.34179	0.72936	QCIF
Akiyo	47.40	47.10	1.5433	1.8054	QCIF
Suzie	44.2	43.5	1.80594	1.9103	QCIF
Claire	45.6	45.3	0.874	1.205	CIF

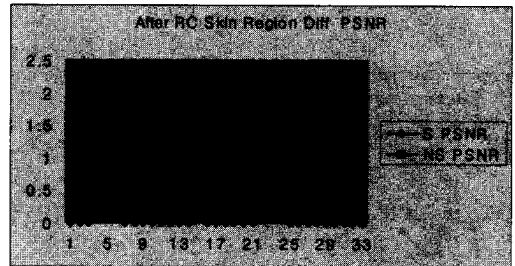
과는 제어전과 후의 얼굴 영역에서 나타나는 PSNR의 증분으로 전경에 해당하는 얼굴 영역의 화질은 배경이나 주변의 화질에 비해 다소 양호하게 나타남을 확인할 수 있었다. 하지만 우선권이 낮은 영역이 높은 영역에 비해 울 제어 후의 화질이 매우 저하될 것으로 생각이 되나, 사실은 울 제어 과정에서 영역의 크기에 따라 영역별로 발생한 비트량이 비례하므로 화질 저하는 그다지 심하지 않았다. 부득이 화질 저하가 심화될 경우, 제안한 기법이 가져다 주는 효과는 프레임 레벨 단에서 프레임을 스킵하므로써 우선권이 높은 영역의 화질은 다소 보호된다. 또한, 프레임에 배정된 타겟 비트가 너무 작을 경우에는 QP=31로 양자화해도 타겟 비트를 넘어서는 경우가 발생하는데 이 경우에는 다음에 부호화될 프레임의 버퍼 자원을 사용하게 되어 다음 프레임에 약간의 영향을 미치는 결과를 초래한다.



(a) 울 제어 결과
(타겟 비트 : 4000, 0~50 Frame)



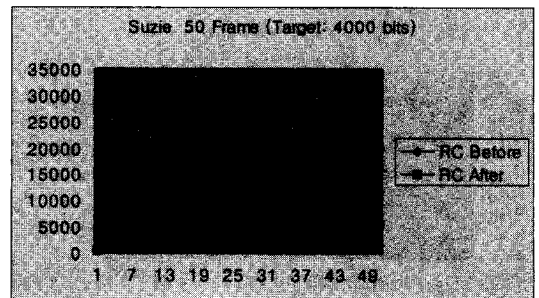
(b) 울 제어 전 후 PSNR



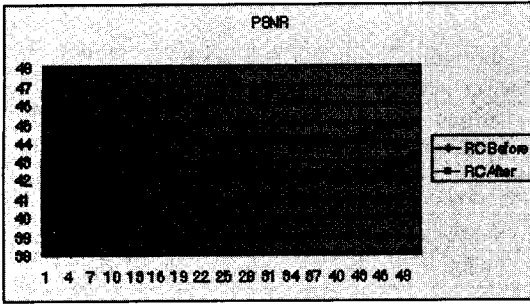
(c) 살색 영역과 비 살색 영역의 울 제어 전후의 PSNR 비교

그림 12. Akiyo 이미지

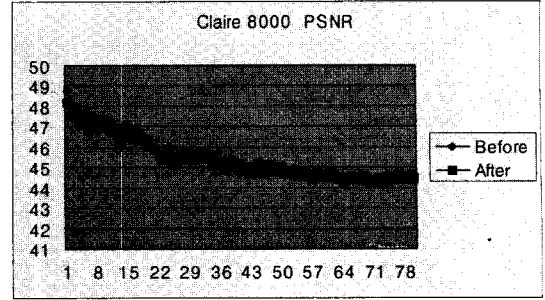
그림 13(c)를 보면 울 제어 후의 살색 영역의 PSNR 차가 25 프레임부터는 비 살색 영역보다 커지는 결과가 초래되었는데, 이는 Suzie 영상이 가지고 있는 특징으로서 QCIF 화면의 1/2이 살색 영역에 해당하는 얼굴 영역이며 또한 얼굴 영역의 움직임이 전체적으로 커 프레임 수가 늘어날 수록 누적 에러가 복원되지 않는 결과를 초래했다. 이는 프레임 계층 울 제어 단에서 적절하게 프레임을 스킵하면 이러한 화질 저하 문제를 방지할 수가 있다. 또한 지나치게 타겟 비트를 낮게 잡아도 이러한 문제가 발생하므로 제어 시스템에서는 이러한 특성을 고려해서 설계를 해야한다.



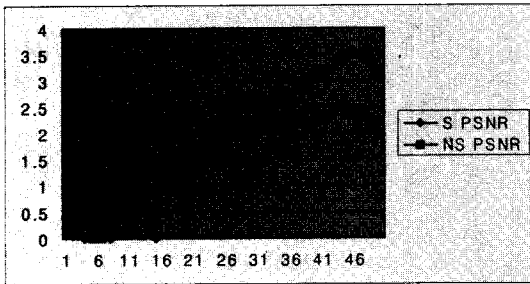
(a) 울 제어결과
(타겟비트 : 4000, 0~50 Frame)



(b) 울 제어 전후 PSNR

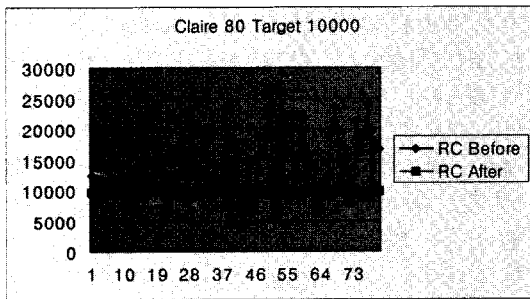


(c) 울 제어 전 후 PSNR (타겟 비트 8000)

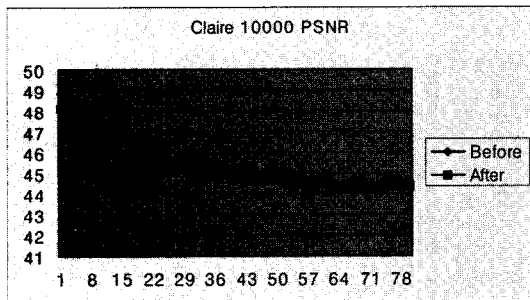


(c) 살색 영역과 비 살색 영역의 울 제어 전후의 PSNR 비교

그림 13. Suzie 이미지



(a) 울 제어 결과
(타겟 비트 : 10000, 0~80 Frame)



(b) 울 제어 전 후 PSNR
(타겟 비트 : 10000)

그림 14. Claire 이미지(CIF : 0~80 Frame) 울 제어 결과

V. 결론

전송 대역이 매우 한정된 채널에 영상 정보를 전송할 경우에는 영상의 원 화질 유지보다는 필요한 정보만을 전송하여 부호화 효율을 높여야만 한다. 이에 대한 대표적인 응용으로 MPEG4의 가상 미팅 (Virtual Meeting) 응용에서는 Face Animation 기법이 이용되는데 이는 FAP(Facial Animation Parameter)를 이용해 2kbps까지 정보를 압축한다.^[9] 마찬가지로 제한된 채널 상에서 영상 통신을 할 경우에는 영상 전체의 화질 유지보다는 중요한 부분만을 전송하여 주관적 화질을 유지함으로써 부호화 효율을 높일 수가 있다. 본 논문에서 제안한 울 제어 방법은 이러한 개념을 응용한 것으로서 낮은 비트율에서 영상의 전경에 해당하는 얼굴 영역의 화질을 보호해 주관적인 성능을 향상시킬 수가 있었다. 또한 이 기법은 사용자에 의해 우선권이 주어지는 영상 객체나 영상의 일정 영역에 대해서도 적용 가능하다. 앞으로 보안해야 할 부분은 영역 분할 시 패닝(panning)이나 카메라의 손떨림이 있는 영상에 대한 적절한 보상 병합 방법과 울 제어 후 복원된 영상에서 발생하는 영역별 블록화 현상을 보상할 후처리 필터의 구현이 필요하다.

향후 멀티미디어가 지향하는 서비스는 사용자의 환경을 보다 친숙하고 강력하게 만들며 다양한 미디어를 지원하고 영상의 경우에는 프로세서의 성능 향상에 따라 고등 알고리즘을 적용하며 영상에 의미를 부여하는 환경으로 진화하고 있다. 영상에 의미를 부여하는 과정은 많은 처리 시간과 인식이라는 복잡한 과정이 필요하나 동일한 환경에서 선택적인 표현이 가능하고 사용자의 편집 기능 및 조작을 가능케 한다. 본 논문에서 제안한 기법들은 이러한 추세에 다소 근접한 것으로서 향후 실시간 화자

인식, 얼굴영역 및 의미 영역 인식기법의 적용이 필요한 시스템의 경우에는 제안한 기법들을 기반으로 응용이 가능하리라 생각한다.

참 고 문 헌

[1] K.R RAO, JJ HWANG "Techniques and Standards for Image, Video, and Audio Coding", *Prentice Hall* 1997.

[2] DRAFT H.263, "Video Coding for Low Bit Rate Communication", 27 January ,1998

[3] J.I Ronda, M.Eckert, S.Rieke, F.Jaureguizar "Advanced Rate Control for MPEG-4 Coders" pp.383-394, *Visual Communication and Image Processing '98*

[4] Jungwoo Lee, "New MPEG-2 Rate Control Algorithms Based on Motion Estimation", pp. 395-404, *Visual Communication and Image Processing '98*

[5] ITU-T Video Codec Test Model, Near-Term, Version8(TMN8)

[6] 이봉호, 서정구, 광노윤, 강태하, 황병원, "영역 분할 정보를 이용한 고속 움직임 추정 부호화 및 의미 영역 인식" pp.665-668, *Proceedings of IEK Summer Conference '98*

[7] Demas Snager, Takuya Asada, Hiedeki Haneishi and Yoichi Muyake, "Facial Pattern Detection and Its Preferred Color Reproduction", pp. 149-152, *The 2nd IS&T/Sid Color Imaging Conference*

[8] Limin et al, "Bit Allocation for Joint Coding of Multiple Video Programs", pp.149-158, *Visual Communication and Image Processing '98*

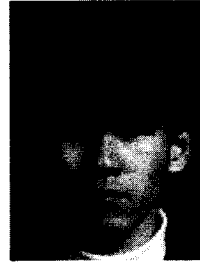
[9] ISO/IEC JTC1/SC29/WG11/N2322, "MPEG-4 Application Document"

[10] Milan Sonka, Vaclav Hlavac, Roger Boyle, "Image Processing, Analysis and Machine Vision". *Chapman & Hall Computing*.

[11] Demas Sanger, Takuya Asada, Hiedeki Haneishi and Yoichi Muyake, "Facial Pattern Detection and Its Preferred Color Reproduction", *The 2nd IS&T/Sid Color Imaging conference*, pp. 149-152, 1994

이 봉 호(Bong-Ho Lee)

정회원



1997년 2월 : 한국항공대학교 전자공학과 졸업(학사)
 1999년 2월 : 한국항공대학교 전자공학과 석사과정 졸업(공학석사)
 1999년 7월~현재 : 한국전자통신연구원(ETRI) 연구원

<주관심 분야> 영상신호처리, 디지털 방송, MPEG-4, 멀티미디어 시스템

전 경 재(Kyung-Jae Jeoun)

정회원

1991. 충남대학교 전산학
 1991. 1~현재 : 한국전자통신연구원(ETRI) 선임연구원

<주관심 분야> 인터넷 프로토콜, 위성 멀티미디어 시스템

광 노 윤(No-Yoon Kwak)

정회원



1994년 2월 : 한국항공대학교 항공전자공학과 졸업(공학사).
 1996년 2월 : 한국항공대학교 대학원 항공전자공학과(공학석사).

2000년 2월 : 한국항공대학교 대학원 항공전자공학과(공학박사)

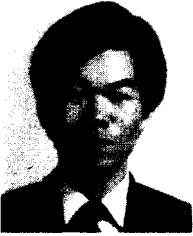
1998년 3월~2000년 2월 : 유한대학 정보통신과 겸임 교수

2000년 3월~현재 : 천안대학교 정보통신학부 전임강사

<주관심 분야> 정지영상 및 동영상 압축, 칼라 영상 처리 및 검색, 컴퓨터 비전, 영상 디스플레이

강 태 하(Tae-Ha Kang)

정회원



1985년 2월 : 한국항공대학교
항공통신공학과 졸업
(공학사)

1987년 2월 : 한국항공대학교
대학원 항공전자공학과
(공학석사)

1995년 3월~현재 : 한국항공대학교 대학원 항공전자공학과(박사과정)

<주관심 분야> 영상 디스플레이, 패턴인식, 동영상 신호처리

황 병 원(Byong-Won Hwang)

정회원



1972년 2월 : 한국항공대학
항공전자공학과 졸업
(학사).

1981년 3월 : 동경대학 대학원
전기공학과(공학석사).

1984년 3월 : 동경대학 대학원
전기공학과(공학박사).

1984년 5월~1985년 3월 : 국방과학연구소 선임연구원.

1993년 8월~1994년 8월 : 동경대학 전기공학과 교환교수.

1985년 8월~현재 : 한국항공대학교 항공전자공학과 교수.

<주관심 분야> 영상처리 및 영상인식, 동영상 신호처리, 교통류 제어