

전역/지역 움직임 정보를 이용한 선택적 부호화 기법

正會員 李 鍾 培*, 金 成 大**

Selective Coding Scheme Using Global/Local Motion Information

Jong Bae Lee*, Seong Dae Kim** *Regular Members*

요 약

본 논문에서 제안한 선택적 부호화 기법은 배경(background)과 목표물(target region) 간에 비트 할당을 차별화하여 일련의 영상을 부호화하는 방식이다. 본 방식은 먼저 전역이동 파라미터(global motion parameter)와 지역이동벡터(local motion vector)를 추출하고, 다음에 계층적 클러스터링 기법(clustering technique)과 quadtree 기법을 사용하여 배경과 목표물을 분할한다. 최종적으로 분할된 배경과 목표물에 대하여 양자화 비트를 차별화함으로써 원하는 화질로 복원할 수 있도록 한다. 본 논문에서는 시뮬레이션을 통하여 제안한 방식이 기존의 방식에 비하여, 특히 배경이 변하고 목표물이 배경보다 충분히 작을 경우에 성능이 우수함을 보인다.

ABSTRACT

A selective coding scheme is proposed that describes a method for coding image sequences distinguishing bits between background and target region. The suggested method initially estimates global motion parameters and local motion vectors. Then segmentation is performed with a hierarchical clustering scheme and a quadtree algorithm in order to divide the processing image into the background and target region. Finally image coding is done by assigning more bits to the target region and less bits to background so that the target region may be reconstructed with high quality. Simulations show that the suggested algorithm performs well especially in the circumstances where background changes and target region is small enough compared with that of background.

I. 서 론

본 논문에서 다루고자 하는 전술영상에서는 일반

적으로 전송채널용량은 제한되어있고, 카메라의 움직임으로 인하여 영상전체가 일정한 속도를 갖는다. 또, 이러한 전술영상에서는 목표물 영역이 영상 전체에서 작은 부분에 위치하지만 그 부분의 정보의 가치는 대단히 크다. 이러한 상황 하에서는 전송채널용량이 제한되어 있으므로 초저속 영상 압축기법(very low bit rate coding algorithm)이 요구되며, 카메라의 움직

*LG 電子 (LG Electronics Co.)

**韓國科學技術院(KAIST)

論文番號: 94305-1103

接受日字: 1994年 11月 3日

임으로 인하여 영상전체가 일정한 속도를 갖고 있으므로 전역이동 보상이 행해져야 한다. 특히 목표물 영역이 영상 전체에서 국부적으로 위치하며 배경 영역에 비해 중요도가 크므로, 영상 부호화시 목표물을 고화질로 복원할 수 있도록 해야 한다.

그런데 기존의 H.261[1]기법이나 이를 수정하여 8kbps, 16kbps 등의 초저속 압축·전송에 응용가능하도록 한 기법(예를 들어 SIM3[2])들은 본 논문에서 다루고자 하는 영상에는 적용이 힘들다. 그 이유를 살펴보면 다음과 같다. 먼저 단순한 BMA(block matching algorithm)를 이용하기 때문에 실제 움직임이 회전(rotation)이나 zooming을 포함한 경우 오정합(false matching)이 발생한다. 즉 BMA는 운동체(rigid body)이며 병진운동(translational motion)만을 한다는 가정하에서 x와 y 방향의 병진운동만을 추정하는 기법이므로 카메라의 움직임으로 인하여 운동체의 회전, zoom 등이 발생한 경우 '단일강체의 병진운동' 조건만으로는 정확한 이동벡터를 구하기 힘들다. 다음에 매크로 블록(16×16)이나 블록(8×8) 단위로 속도벡터를 측정하여 전송하기 때문에 많은 속도벡터 전송비트가 필요하게 된다. 즉 속도벡터를 전송하는 정보도 초저속 압축·전송에서는 상당한 비중을 차지하기 때문에 가능한한 같은 속도벡터를 갖는 영역을 합쳐서 이들 대표속도를 구하고 이를 전송하는 것이 유용한 경우가 많다. 전송영상에서는 목표물 영역이 배경영역에 비하여 정보의 중요성이 크다. 그런데, 기존의 방식에서는 배경과 목표물 영역 간에 정보의 중요도를 고려하지 않고, 같은 부호화 기법을 사용하기 때문에 목표물을 고화질로 복원하기가 쉽지 않다. 즉 배경에 비해 훨씬 중요한 영역인 목표물 영역이 배경과 동일하게 취급되어서 정보량이 훨씬 작은 배경과 같은 화질로 부호화되는 단점이 있다. 이러한 기존의 부호화 방식의 여러 문제점들을 해결하기 위하여 본 논문에서는 전역/지역이동 파라미터를 이용하여 목표물과 배경을 분할하고 목표물에 충분한 비트를 할당하여 영상 부호화를 행함으로써 목표물을 고화질로 복원할 수 있는 방식을 제안한다.

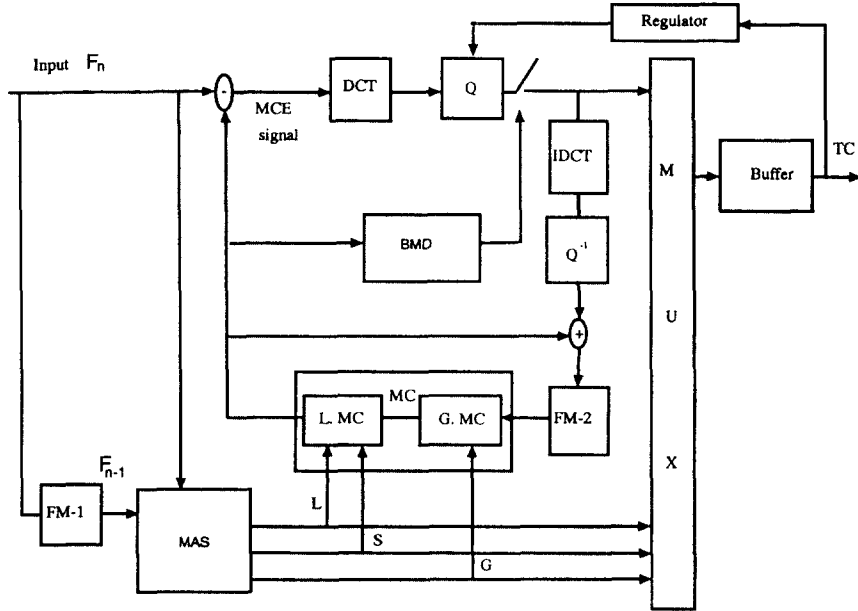
본 방식은 먼저 회전, zoom등의 카메라의 움직임을 보상하기 위하여 네개의 전역이동 파라미터를 추출한다. 이렇게 전역이동 파라미터가 추출되면 BMA를 이용하여 현재영상과 전역이동을 보정한 과거영

상 간에 다시 지역이동 파라미터를 추출하고, 이들 속도값으로부터 배경영역과 목표물 영역을 분할한다. 분할과정은 먼저 계층적 클러스터링(hierarchical clustering) 방법을 이용하여 초기 배경영역을 결정하고, 나머지 영역에 대해 quadtree 기법으로 목표물 영역과 배경영역을 세밀히 분할한다. 최종적으로 목표물 영역에 충분히 양자화 비트를 할당하고 배경 영역에는 그 나머지 비트를 할당하여 영상 부호화를 행함으로써 목표물 영역을 고화질로 복원할 수 있도록 한다.

본 논문은 다음과 같이 구성되었다. 제 2장에서는 제안한 방식의 개략적인 구조를 나타내고 있으며 제 3장에서는 기존의 BMA 방식으로부터 추출한 국부 이동벡터 정보로부터 네개의 전역이동 파라미터를 추출한 방식을 설명하고 있다. 또 제 4장에서는 계층적 클러스터링 방식과 quadtree 기법을 이용하여 배경과 목표물을 분할하는 방식을 제시하고 있으며 제 5장에서는 분할된 배경과 목표물 영역에 대하여 양자화 비트를 차별화 하여 데이터를 전송하는 방식을 제시한다. 마지막으로 제 6장에서는 시뮬레이션을 통하여 본 방식의 우수성을 입증한다.

II. 본 방식의 개략적인 구조

그림 1은 제안한 방식의 블록 다이어그램을 나타내고 있다. 먼저 Motion analysis & Segmentation(MAS) 블록은 현재영상 F_n 과 과거영상 F_{n-1} 의 데이터를 입력으로 받아들여서, 지역이동 파라미터(L), 전역이동 파라미터(G), 분할정보(Segmentation Information: S)를 계산하여 출력으로 내보낸다. 프레임 메모리 1(FM-1)와 프레임 메모리 2(FM-2)는 각각 원래의 이전영상과 각종 파라미터와 오차신호로부터 복원된 이전영상을 저장하고 있다. MAS에서 출력된 각종 파라미터들은 기존의 DPCM을 이용하여 예측 부호화되고, 전송채널(Transmission Channel: TC)로 출력되며, 동시에 움직임 보상(MC)블록으로 피드백(feed-back)된다. MC블록은 다시 전역이동 보상을 행하기 위한 전역 MC(G, MC) 블록과 지역이동 보상을 행하기 위한 지역 MC(L, MC) 블록으로 세분화된다. 목표물 영역에서는 움직임보상오차(MCE) 신호는 DCT 및 양자화(Q) 블록에 의하여 부호화되고 이들이 전송채널로 출



FM : Frame Memory, MAS : Motion Analysis and Segmentation
 L : Local Motion Parameters, S : Segmentation Information
 G : Global Motion Parameters , MC : Motion Compensation
 FM : Frame Memory, TC : Transmission Channel

그림 1. 제안한 방식의 블록 다이어그램
 Fig. 1 Block Diagram of Suggested Method

력되며 동시에 역양자화(Q^{-1}) 블록을 거쳐 움직임 보상된 이전영상에 더해져서 FM-2에 저장된다. 배경영역의 경우에는 배경 전송 결정(BTD) 블록에 의해 주어진 값보다 MCE 신호가 큰 경우에만 스위치(SW-1)을 닫아서 전송 채널로 출력하고 동시에 FM-2의 입력으로 이용될 수 있도록 한다. 또, 배경영역에서는 조절기(regulator)를 통해 조절해줌으로써, 전송채널 용량을 초과하지 않도록 해준다.

Ⅲ. 전역이동 파라미터 예측

블록정합기법(BMA)[3]은 각 블록 단위의 병진운동 성분 밖에는 추정하지 못하기 때문에 여러 이동성분이 복합적으로 결합된 실제영상에서는 오정합(false matching)이 발생하여 정확한 이동벡터 필드(motion vec-

tor field; MVF)를 추정하지 못하는 경우가 생긴다. 특히 회전, zoom 등을 포함한 전역이동을 추정하는데 있어서 오정합이 심하게 일어나며, 이러한 이유로 서론에서 언급한 바와 같이 H.261[1], SIM3[2] 등의 알고리즘이 제대로 적용되지 않는다. 이렇게 BMA에 의하여 잘못 추정된 이동벡터를 영상 부호화에 이용하면 예측 후 재생된 영상의 주관적인 화질저하, 이동벡터에 대한 손송데이터의 증가, 보정된 영상의 화질저하 등을 일으키는 단점이 있다. 이러한 이유로 전역이동 파라미터 예측에 관하여 지금까지 여러 연구가 행해져 왔다[4-7]. Hötter[4]는 시간-공간 경사 방식(Spatio-temporal gradient method)을 이용하여 움직임보상 오차신호(motion-compensated error signal)의 자승을 최소화함으로써 전역이동 파라미터를 추출하였다. 마찬가지로 Wu와 Kittler[5]는 영상의 해상도

(resolution)를 달리하면서 네개의 전역이동 파라미터를 구하는 미분추정방식(differential estimation method)을 제안하였다. 그러나 이러한 미분추정방식은 시간 공간 경사방식을 사용하므로 밝기정보가 얼마나 잘 개략화할 수 있는가에 따라 성능이 결정된다. 그러므로 움직임이 크게 일어난 경우, 위 방식들은 여러 번의 반복과정(iteration process)이 필요하며, 관찰점(observation point)이 충분하지 않을 경우 발산할 우려가 있다. 한편 Adolph와 Buschmann[6]는 프레임 정합에 의하여 전역이동 예측방식을 제안하였다. BMA에서 2개의 이동성분으로서 움직임을 추정하는 것과는 달리 그는 3 개의 파라미터를 이용하여 프레임 정합을 행하였는데, 정합 척도(matching criterion)로는 MAD(mean absolute difference)를 사용하였다. 또, Tse와 Baker[7]는 기존의 BMA에 의해 얻어진 지역 이동벡터를 이용하여 3 파라미터에 의하여 카메라 zoom과 pan을 보상하는 방식을 채택하였다. 본 방식에서는 전탐색 BMA(full search BMA)에 의해 얻어진 지역 이동벡터를 이용하여 네개의 전역이동 파라미터를 구하는 방식을 제안하고자 한다. 이 방식은 기존의 방식[7-11]과 유사하지만 카메라의 회전효과(rotation effect)를 잘 표현할 수 있는데, 그 이유는 회전각이 매우 작지 않는 한, 3 파라미터로 회전의 영향을 제대로 보상할 수 없기 때문이다. 따라서 본 논문에서는 회전, zoom 등의 다양한 움직임을 보정해 주기 위하여 4 파라미터를 이용하여 전역이동을 예측하는 방식을 제안하고자 한다. 일반적으로 우리가 다루고자 하는 전술영상에서는 목표물 영역은 배경영역에 비하여 매우 작고, 전송채널용량은 매우 제한되어 있다. 따라서 본 연구에서 다루고자 하는 영상은 아래와 같은 특징을 만족한다고 가정한다.

〈가 정〉

1. 배경은 카메라의 움직임에 인하여 전체적인 움직임을 갖고, 배경의 속도는 목표물의 속도와는 다른 값을 갖는다.
2. 목표물 영역은 배경영역에 비하여 충분히 작다.

먼저 전역이동 파라미터를 예측하기 위해 몇가지 값들을 정의하자. (X, Y) 와 (X', Y') 는 각각 카메라의 움직임이 일어나기 전과 일어난 후의 영상공간 좌표

라 하고 F_z 를 zoom factor, P_x, P_y 를 각각 x와 y 방향의 panning 파라미터라 가정하자. 또, x와 y 방향의 회전중심(rotation center)을 각각 g_x, g_y 라 하면, X', Y' 는 각각 식 (1)과 같이 표시된다.

$$\begin{aligned} X' &= F_z[(X - g_x)\cos(\theta) - (Y - g_y)\sin(\theta)] + g_x + P_x, \\ Y' &= F_z[(X - g_x)\sin(\theta) + (Y - g_y)\cos(\theta)] + g_y + P_y. \end{aligned} \quad (1)$$

여기서 θ 가 충분히 작을 경우, $\cos(\theta)=1, \sin(\theta)=\theta$ 이 되며, 회전중심 g_x, g_y 를 원점으로 놓을 경우 식 (1)은 식 (2)와 같이 된다.

$$X' = F_z(X - Y\theta) + P_x, Y' = F_z(X\theta + Y) + P_y \quad (2)$$

또, $f_z = a_1, f_z\theta = a_2, P_x = a_3, P_y = a_4$ 로 놓으면 식 (2)는 다시 식 (3)과 같이 표시된다.

$$X' = a_1 X - a_2 Y + a_3, Y' = a_2 X + a_1 Y + a_4, \quad (3)$$

미찬가지로 X, Y 는 식 (4)와 같이 표현되며, 이들로부터 예측된 속도벡터는 식 (5)와 같다.

$$X = a_1' X' - a_2' Y' + a_3', Y = a_2' X' + a_1' Y' + a_4' \quad (4)$$

여기서 $a_1' = a_1/(a_1^2 + a_2^2), a_2' = a_2/(a_1^2 + a_2^2), a_3' = 1/(a_1^2 + a_2^2)[-a_1 a_3 + a_2 a_4], a_4' = 1/(a_1^2 + a_2^2)[a_2 a_3 + a_1 a_4]$ 이다.

$$\begin{aligned} \hat{u}_{ij} &= X_i - X'_i = (a_1' - 1)X'_i + a_2' Y'_j + a_3' \\ &= b_1 X'_i + b_2 Y'_j + b_3, \\ \hat{v}_{ij} &= Y_j - Y'_j = -a_2' X'_i + (a_1' - 1)Y'_j + a_4' \\ &= -b_2 X'_i + b_1 Y'_j + b_4 \end{aligned} \quad (5)$$

단 $b_1 = a_1' - 1, b_2 = a_2', b_3 = a_3', b_4 = a_4'$ 이다.

다음에 각 블록의 국부속도벡터(local velocity vector) $W_{ij} = (u_{ij}, v_{ij})^T$ 를 $N \times N$ 창틀에 대하여 구하며, 창틀의 크기는 영상의 특성과 적용분야에 따라 결정하는데 통상 $N \times N$ 으로 한다. 전역이동 파라미터를 예측하기 위해서 속도벡터 오차의 l_2 norm $E(b_1, b_2, b_3, b_4)$ 을 식 (6)과 같이 정의하자.

$$E(b_1, b_2, b_3, b_4) = \sum_{X_i, Y_j \in M} \left\| (b_1 X_i' + b_2 Y_j' + b_3 - u_{i,j})^2 + (b_2 X_i' + b_1 Y_j' + b_4 - v_{i,j})^2 \right\| \quad (6)$$

여기서 M은 각 창들의 중심점 들의 집합이다. 다음에 각 파라미터에 대하여 식 (6)을 미분하여 그 값을 영으로 둬으로써 최적 전역이동 파라미터를 아래와 같이 구할 수 있다.
 즉 $\partial E / \partial b_k = 0, (k=1, 2, 3, 4)$ 이다.

$$B = S^{-1} U \quad (7)$$

$$S = \begin{pmatrix} \sum(X_i'^2 + Y_j'^2) & 0 & \sum X_i' & \sum Y_j' \\ 0 & \sum(X_i'^2 + Y_j'^2) & \sum Y_j' - \sum X_i' & 0 \\ \sum X_i' & \sum Y_j' & \sum 1 & 0 \\ \sum X_i' & \sum Y_j' & 0 & \sum 1 \end{pmatrix} \quad (8)$$

$$B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} \quad (9)$$

$$U = \begin{pmatrix} \sum(X_i'^2 + Y_j'^2 - u_{ij} X_i' - v_{ij} Y_j') \\ \sum(-u_{ij} Y_j' + v_{ij} X_i') \\ \sum(X_i' - u_{ij}) \\ \sum(Y_j' - v_{ij}) \end{pmatrix} \quad (10)$$

여기서 영상의 중심점을 원점으로 놓으면 S나 U의 항 중에서 X_i' 혹은 Y_j' 의 기수승(odd power)을 포함한 항이 "0"이 되므로 식 (8), 식 (10)은 각각 식 (11), 식 (12)와 같이 간략화된다.

$$S = \begin{pmatrix} \sum(X_i'^2 + Y_j'^2) & 0 & 0 & 0 \\ 0 & \sum(X_i'^2 + Y_j'^2) & 0 & 0 \\ 0 & 0 & \sum 1 & 0 \\ 0 & 0 & 0 & \sum 1 \end{pmatrix} \quad (11)$$

$$U = \begin{pmatrix} \sum(X_i'^2 + Y_j'^2 - u_{ij} X_i' - v_{ij} Y_j') \\ \sum(-u_{ij} Y_j' + v_{ij} X_i') \\ \sum(X_i' - u_{ij}) \\ \sum(Y_j' - v_{ij}) \end{pmatrix} \quad (12)$$

여기서 4 전역이동 파라미터들은 아래 식과 같이 된다.

$$\begin{aligned} a_1 &= \frac{\sum(X_i'^2 + Y_j'^2 - u_{ij} X_i' - v_{ij} Y_j')}{\sum(X_i'^2 + Y_j'^2)}, \\ a_2 &= \frac{\sum(-u_{ij} Y_j' + v_{ij} X_i')}{\sum(X_i'^2 + Y_j'^2)}, \\ a_3 &= \frac{\sum(-u_{ij})}{\sum 1}, \\ a_4 &= \frac{\sum(-v_{ij})}{\sum 1} \end{aligned} \quad (13)$$

복표본 영역(target region)이 전역이동 파라미터에 영향을 주기때문에 측정된 전역이동 파라미터 값과 각 블록의 속도벡터와의 차(difference)를 측정하여, 이 값이 주어진 문턱값 이상이면 이 영역을 제거하고 전역이동 파라미터 예측과정을 2, 3번 반복한다. 최종적으로 전역이동 파라미터가 얻어지면 이 파라미터를 이용하여 과거 영상에 대해 움직임보상을 행하여 움직임 보상된 영상 $U_i(x, y)$ 를 얻는다. 상황에 따라서는 $U_i(x, y)$ 에 대하여 위와 비슷한 과정을 반복함으로써 복표본 영역의 부정역 이동(semi-global motion) 파라미터를 얻을 수 있다.

IV. 영역 분할

1. 초기 배경영역 결정

먼저 움직임 보상된 과거영상 $U_i(x, y)$ 와 현재영상 $C(x, y)$ 에 대하여 full search BMA를 이용하여 각 블록의 속도벡터를 추출한다. 다음에 각 블록 속도벡터의 유사성(similarity)를 측정하고 이로부터 그룹을 정의한다. 즉 각 그룹(group)은 유사한 속도벡터를 갖는 블록들의 집합이다. 본 논문에서는 각 블록에 그룹들을 할당하기 위해서 agglomerative 클러스터링 방법 [12]에 기초한 계층적 클러스터링 기법을 사용하였다.

여기서 각 블록에 그룹 number를 할당하기 위하여 먼저 몇가지 파라미터를 정의한다. 먼저 C_l 를 클러스터 l , P_l 을 C_l 의 부모 클러스터(parent cluster), th 를 속도벡터의 유사성을 판단하기 위한 문턱값, W_{ij} 를 블록 B_{ij} 의 속도벡터, U_i, Z_l 을 각각 같은 클러스터에 속하는 블록들의 평균속도벡터 및 블록의 갯수로 놓자.

또, 주어진 영상에 대하여 다음과 같이 가정한다.

- 가정 1. 주어진 영상이 $M \times N$ 블록으로 구성되어 있다.
- 가정 2. $i=0$ 일 경우 $U_{(i-1)j}$ 를 U_{ij} 로, $V_{(i-1)j}$ 를 V_{ij} 로 놓고, $j=0$ 일 경우 $U_{i(j-1)}$ 를 U_{ij} 로, $V_{i(j-1)}$ 를 V_{ij} 로 놓는다.
- 가정 3. $B_{\{ij\}}$ 가 클러스터 C_l , $B_{(i-1)j}$ 이 클러스터 $B_{(i-1)j}$, $B_{i(j-1)}$ 이 클러스터 C_m 에 속한다.

위에서 정의한 파라미터와 가정에 따라 아래와 같이 그룹 알고리즘을 제안한다.

(그룹 number 할당)

Step 1. 각종 파라미터를 초기화 한다. 먼저 $l=0, j=0, k=0, m=0$ 으로 놓는다. 다음에 블록 b_{ij} 가 클러스터 C_l 에 속한다고 가정한다. 즉 $C_k = b_{ij}$. 또 P_l 과 C_l 을 같게두고 U_m, U_n 의 초기값을 U_{ij} 로 놓는다.

Step 2. 먼저 i 를 증가시킨다. 만약 i 가 M 일 경우에는 j 를 증가시키고, $i=0$ 으로 놓는다. 만약 $j=N$ 이면 step 4로 간다.

Step 3. 아래의 네가지 경우에 따라 현재블럭이 어느 클러스터에 속하는 지를 결정한다.

Case 1: $\|U_{ij} - \bar{U}_l\| < th$ 이고, $\|U_{ij} - \bar{U}_m\| < th$ 이면 $(Z_l \bar{U}_l + U_{ij}) / (Z_l + 1)$ 을 새로운 \bar{U}_l 로, P_l 을 P_m 으로 놓는다. 또, Z_l 을 1 증가시키며, 블록 B_{ij} 를 클러스터 C_l 에 포함시킨다.

Case 2: $\|U_{ij} - \bar{U}_l\| < th$ 이고, $\|U_{ij} - \bar{U}_m\| > th$ 이면 $(Z_l \bar{U}_l + U_{ij}) / (Z_l + 1)$ 을 새로운 \bar{U}_l 로 놓는다. 또, Z_l 을 1 증가시키며 블록 B_{ij} 를 클러스터 C_l 에 포함시킨다.

Case 3: $\|U_{ij} - \bar{U}_l\| > th$ 이고, $\|U_{ij} - \bar{U}_m\| < th$ 이면 $(Z_m \bar{U}_m + U_{ij}) / (Z_m + 1)$ 을 새로운 \bar{U}_m 로 놓는다. 또, Z_m 을 1 증가시키며 블록 B_{ij} 를 클러스터 C_m 에 포함시킨다.

Case 4: $\|U_{ij} - \bar{U}_l\| > th$ 이고, $\|U_{ij} - \bar{U}_m\| > th$ 이면 $(Z_m \bar{U}_m + U_{ij}) / (Z_m + 1)$ 을 새로운 \bar{U}_m 로 놓는다. k 를 1 증가시키고 B_{ij} 를 C_k 로 포함시킨다. 또 Z_k 를 1로 두고 P_k 에 k 를 할당하며, U_{ij} 를 새로운 \bar{U}_l 으로 놓는다.

Step 4. 부모 클러스터의 값이 같은 블록을 같은 클

러스터로 놓고 클러스터 갯수 및 속도벡터를 갱신한다. 즉 $k_1 < k_2$ 이고 $P_{k_1} = P_{k_2}$, $Z_{k_1} = a_1$, $Z_{k_2} = a_2$ 인 경우 새로운 클러스터 C_{k_3} 는 C_{k_1} 과 C_{k_2} 를 합친값이 되며, $Z_{k_3} = Z_{k_1} + Z_{k_2}$, 이 된다.

Step 5. 클러스터 갯수가 큰 순서대로 재배열(sorting)을 행하여 각 블록에 대한 최종 클러스터의 값을 결정한다.

이와같이 각 블록에 대하여 클러스터 그룹이 결정되면, 초기 배경영역을 구하게 된다. 이는 가정에 의해서 목표물 영역이 배경영역에 비해 충분히 작으므로 가장 큰 그룹이 초기 배경영역이 된다.

2. Quad-tree [13]를 이용한 목표물 영역의 세분화

앞에서 언급한 것처럼 BMA 수행시 블록 크기를 고정시킬 경우 여러가지 문제점이 있다. 따라서 quadtree를 이용하여 블록 크기를 가변적으로 조정하므로써 계산량이 증대되는 단점은 있지만 부가 정보량과 예측이득 등의 부호화 효율이 우수하기 때문에 여러 연구가 행해졌다[14-19]. 일례로서 P.Stobach[14, 15]는 움직임 보상 영상에 대하여 quadtree 기법을 이용하여 영상분할을 행하였다. 그는 먼저 이동보상 오차영상에서 16×16 DCT 짐중효율이 30~40% 정도에서는 오히려 DCT 하기 전보다 떨어짐을 보였다. 이에 따라 DCT 변환영상 대신에 원래의 이동보상영상(motion compensated image)에 대하여 오차블럭의 균질성(homogeneity)을 조사하고, 균질성을 만족하는 경우 그 블럭에 대한 quadtree 과정을 끝낸다. 그렇지 않으면 네개의 서브블럭으로 나눈 다음, 각 서브블럭에 대하여 같은 과정을 반복한다. 이렇게 가변블럭으로 분할되면 DCPM을 이용하여 영상 부호화를 행하였다.

본 논문에서는 영상의 오차신호와 속도정보를 동시에 고려하여 초기 배경영역으로 정의되지 않는 부분을 quadtree 방식으로 분할하였다. 본 quadtree 방식은 top-down 방식으로, 주어진 블럭이 하나의 leaf로 정의되거나 네개의 서브블럭(sub block)으로 나뉘어진다. 각 블럭이 분할되면, 마찬가지로 과정으로 해당 블럭의 서브블럭에 대해 보다 세밀한 영역 분할이 필요한 지를 결정하게된다. 여기서 가장 효과적인 영상 압축 알고리즘은 비트율(bit rate)과 왜곡(distortion)

을 어떻게 효과적으로 줄이느냐 하는데 있다. 이는 R-D 평면에서의 최적점을 선택하는 문제이다. 즉 Lagrange multiplier($\lambda \geq 0$)를 이용하여 각 영역 k에 대하여 아래와 같은 목적함수(objective function)를 최소화함으로써 R-D 평면에서의 최적점(optimal point)을 구할 수 있다.

$$\min \{d_k + \lambda b_k\} \tag{14}$$

여기서 b_k 는 각 영역에 할당된 비트의 수, d_k 는 이와같이 비트를 할당하였을 경우 발생하는 왜곡, λ 는 Lagrange multiplier를 나타낸다.

이를 'principle of separate minimization'이라 부르는 데, 그 이유는 Shoham과 Gersho[18]가 각각의 분리된 영역에서의 목적함수를 최소화함으로써 전체영역의 최적화가 가능하다는 것을 보였기 때문이다. 일례로써 크기 $2^{n+1} \times 2^{n+1}$ 의 블록 X_{n+1} 의 최적화 quadtree 구조를 구하는 방법에 대하여 생각해보자. 먼저 $2^n \times 2^n$ 의 네 서브블럭에 대한 최적 quadtree 구조를 알고있다고 가정하고, 각 서브블럭의 비트율과 왜곡을 각각 $b_n^*, d_n^*, (i=1, \dots, 4)$ 로 놓자. 또 이 네개의 서브블럭을 주어진 블록의 leaf 들로 나타낼 경우 비트율과 왜곡을 각각 b_{n+1}, d_{n+1} 로 놓자. 식 (14)로부터 $C_r \leq 0$ 를 만족하면 네개의 서브블럭은 하나의 leaf로 합쳐진다.

$$C_r = \Delta d - \lambda \Delta b \tag{15}$$

$$\Delta d = d_{n+1} - \sum_{i=1}^4 d_{n,i}^*, \text{ and } \Delta b = \sum_{i=1}^4 b_{n,i}^* - b_{n+1}$$

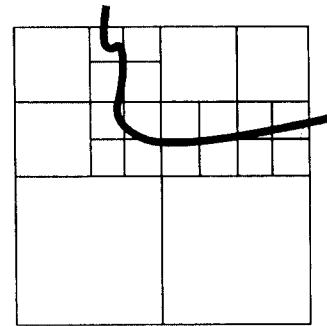
여기서 Δa 와 Δb 는 각각 네개의 서브트리들 하나의 leaf node로 합침으로써 발생하는 왜곡의 증가량과 비트율 감소량을 나타낸다. Quadtree 구조가 그림 2와 같은 경우에 서브트리 X_{n+1} 을 나타내는데 필요한 비트 수는 아래와 같이 표시된다.

$$d_{n+1}^* = \begin{cases} '1' + b_{n+1} & \text{if } \Delta d \leq \lambda \Delta b \\ '1' + \sum_{i=1}^4 b_{n,i}^* & \text{otherwise} \end{cases} \tag{16}$$

그리고

$$d_{n+1}^* = \begin{cases} '1' + d_{n+1} & \text{if } \Delta d \leq \lambda \Delta b \\ '1' + \sum_{i=1}^4 d_{n,i}^* & \text{otherwise} \end{cases}$$

여기서 '1'의 의미는 주어진 블록을 네개의 서브블럭으로 분할할 지의 여부를 나타내는 비트이다.



(a)

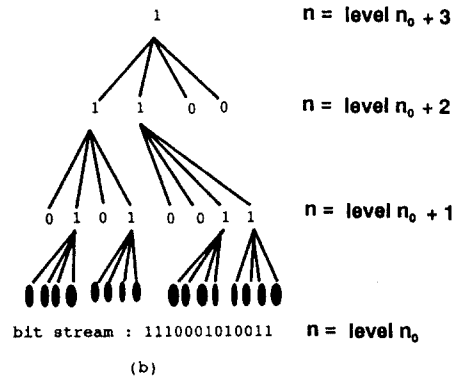


그림 2. Quadtree 데이터 구조
Fig. 2 Quadtree Data Structure

위에서 정의한 왜곡과 비트율을 기본으로 다음과 같은 quadtree 방식을 제안한다.

〈Quadtree 알고리즘을 이용한 영역분할〉

Step 1. 주어진 영상 $M \times N$ 블럭에 대하여 배경으로 정의되지 않는 영역에 대하여 level을 0으로 할당한다.

Step 2. 각 블럭을 서브블럭으로 나누고, 분할된 영

역의 속도벡터를 측정한다. 그리고 level을 1 증가시킨다.

Step 3. 4 분할 영역에 대해서 $C_r \leq 0$ 을 만족하면 각 서브블럭을 합쳐서 하나의 leaf로 놓고 그 블럭에 대한 quadtree 과정을 종료한다. 그렇지 않을 경우 step 4로 간다.

Step 4. 분할된 영역의 그룹 number를 측정한다. 만약 분할된 영역의 그룹 number가 인접한 영역의 그룹 number로부터 구할 수 없으면, 이 영역에 새로운 그룹 number를 할당하고 step 2로 가서 본 과정을 반복한다.

그림 3은 움직임 보상된 과거 영상과 현재 영상간에 quadtree방식을 이용한 영상 분할 방식의 예이다. 영역 I, II, III은 배경을 나타내고 있고 IV, V는 목표물 영역을 나타내고 있다.

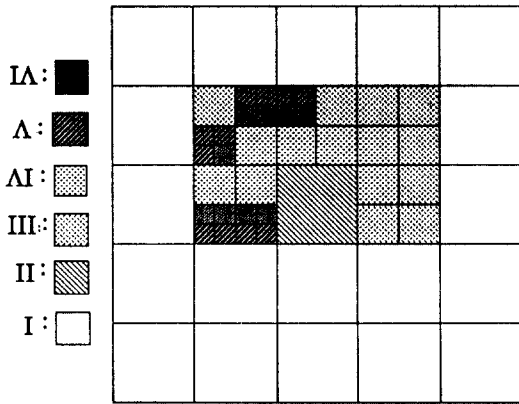


그림 3. Quadtree 방식의 예
Fig. 3 Example of Quadtree Scheme

V. 선택적 부호화와 데이터 전송

배경이 복잡하고 카메라의 움직임에 의해 영상 전체가 일정하게 움직이는 환경에서 전역이동 파라미터를 보상하는 경우에 전 영상에 걸쳐 움직임보상 오차 신호가 존재한다. 우리가 다루고자 하는 전술 영상이 그 대표적인 예로서, 특히 전송 채널의 용량이 한정되어 있어서 고압축률의 영상 압축기법이 요구

된다. 따라서 배경과 목표물 영역 간의 정보의 가치를 고려하지 않고 영상을 압축·전송하는 경우에는 제한된 채널용량으로서 목표물을 고화질(high quality)로 복원할 수 없다. 이를 해결하기 위해서는 배경과 목표물 간의 양자화기 스텝크기(step size)를 다르게 하여야 한다. 먼저 E_{max} 를 목표물 영역에 대한 최대 허용오차로 놓고, 목표물 영역에 속하는 블록 b_i 에 B_i 비트를 할당하였을 경우 $\Theta_T(b_i(B_i))$ 를 목표물 영역에 속하는 블록 b_i 의 양자화 오차신호라 가정하자. 이 경우 재생된 목표물 영역의 화질을 보장하기 위하여 식 (17)이 만족되도록 양자화 비트가 할당되어야 한다.

$$\sum_{b_i \in R_t} \Theta_T(b_i(B_i)) \leq E_{max} \quad (17)$$

여기서 R_t 는 목표물 영역에 속하는 블록의 수이다.

또 R 이 전체 시스템에서 허용하는 최대 비트율이라고 하면 식 (18)을 만족하도록 비트를 할당하여야 한다.

$$B_s \leq \frac{R}{frame\ rate} - \sum_{b_i \in R_t} B_i - P_a \quad (18)$$

여기서 B_s 는 배경신호를 전송하기 위하여 사용될 수 있는 비트율이고, P_a 는 전역이동 파라미터, 그룹 속도 벡터, 위치정보등을 포함하는 파라미터 정보이다.

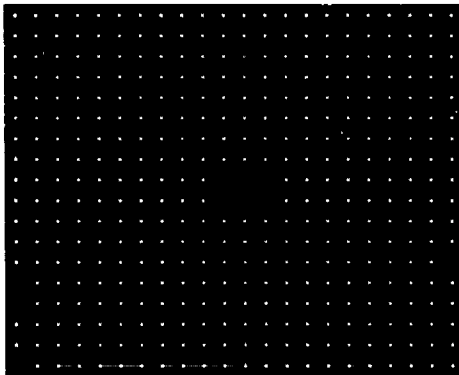
즉 목표물 부분이 우리의 주된 관심분야이므로 배경 부분은 상대적으로 정보의 중요성이 떨어진다. 따라서 본 연구에서는 목표물 영역의 화질을 항상 고정도로 유지할 수 있도록 충분한 비트를 할당하고, 나머지 비트로 배경 부분에 대한 압축·전송을 행하였다.

본 알고리즘이 잘 적용되기 위해서는 목표물 영역이 배경영역에 비하여 충분히 작아야한다. 그 이유는 목표물 영역이 커짐에 따라 목표물 영역의 움직임 보상 오차 신호의 양이 많아지고, 부가적으로 각 목표물 영역에 대한 위치정보, tree level, 그룹 number등의 보조정보가 많이 요구되기 때문이다.

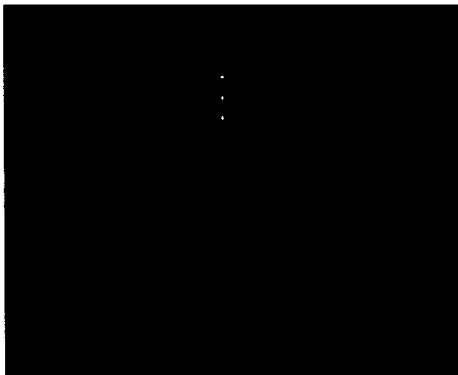
VI. 실험결과 및 결론

본 논문에서는 “Table tennis”영상, “Car 영상”, “Trevor 영상”을 이용하여 시뮬레이션을 행하였다. 먼저 “Table Tennis”의 17번째 프레임과 18번째 프레

임을 각각 이전영상과 현재영상으로 놓자. 제안한 방식에 의하여 예측된 전역이동 파라미터는 $\hat{b}_1=0.078$, $\hat{b}_2=0.035$, $\hat{b}_3=0.001$, 그리고 $\hat{b}_4=0.129$ 값을 갖는데, 이 결과로부터 zoom-out이 전역이동을 결정하는 중요한 요소임을 알 수 있다. 그 이유는 \hat{b}_2 는 zoom factor에 의해 결정되고 \hat{b}_1 , \hat{b}_3 , 그리고 $\hat{b}_4=0.129$ 는 회전 효과(rotation effect)를 포함하는 panning 효과를 결정하기 때문이다. 전역이동을 보상하기 위해서 부화소(subpixel)를 계산하여야 하는데, 이를 위해서 쌍선형 보간(bilinear interpolation)을 사용하였다. 그림 4 (a)와 (b)는 각각 전역이동보상을 행하기 전과 행한



(a) 전역이동 보상 이전
(a) Before global motion compensation



(b) 전역이동 보상 이후
(b) After global motion compensation

그림 4. 예측이동벡터
Fig. 4 Estimated motion vector

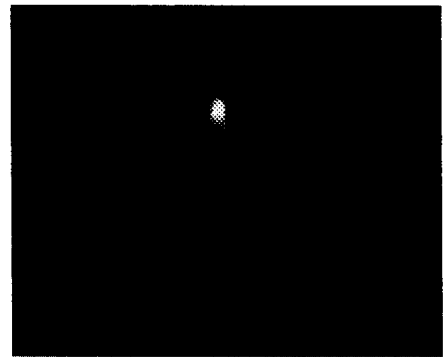
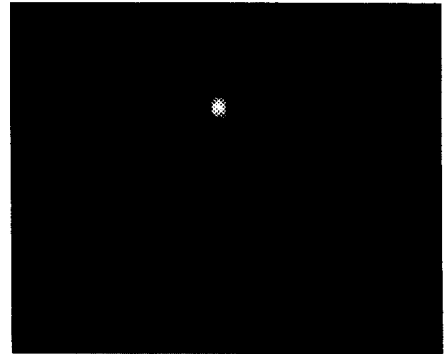


그림 5. 재생된 목표불 영역
Fig. 5 Reconstructed target region

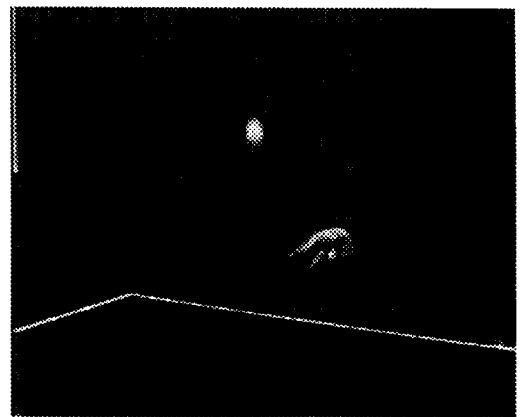


그림 6. 재생영상
Fig. 6 Reconstructed image

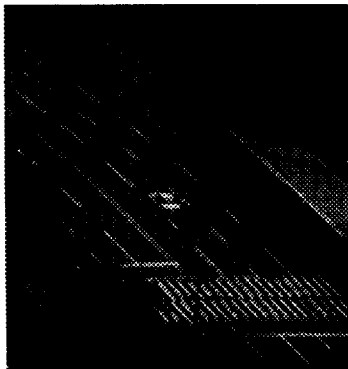
후의 제안한 방식의 움직임 벡터를 표시하고 있다. 이 결과를 통해, 카메라 움직임을 제대로 보상하고, 전송해야할 움직임 벡터를 줄이기 위해 전역이동보상이 필요함을 알 수 있다.

그림 (5), (6)은 각각 제안한 방식으로 재생한 목표물 영역과 영상을 나타내고 있는데, 원영상을 잘 표현하고 있음을 알 수 있다.

특히 목표물 영역으로 설정한 탁구공(ping-pong ball)의 경우, 재생된 영상이 원영상을 고화질로 표현하고 있다. 표 1에서는 제안한 방식(SGFBMA)과 움직임 보상 이전의 전역이동보상(FBMA), 움직임 보상 이후의 전역이동보상(GFBMA) 방식으로 영상의 압축

전송을 행했을 경우의 PSNR 과 비트 수를 나타내고 있으며, 표 2에서는 전송비트의 각 성분들을 나타내고 있다. FBMA의 경우는 재생된 영상의 PSNR이 매우 낮음을 알수있다. 그 이유는 카메라 zoom이나 pan 이 발생하는 영상에서 global motion 보상을 하지 않아서 움직임 보상 오차신호가 영역 전체에 걸쳐 발생하는데, 고압축률의 구현을 위해 오차신호를 따로 보내주지 않았기 때문이다.

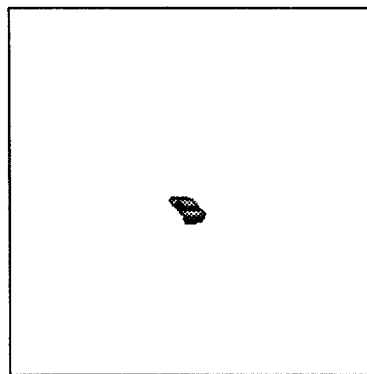
GFBMA의 경우는 FBMA에 비하여 PSNR이 매우 향상되었고, 제안한 방식과 영상전체의 PSNR이 큰 차이가 없음을 알 수 있다. 그러나 목표물 영역에 대해서는 제안한 방식에 비하여 재생성능이 많이 떨어



(a) small car

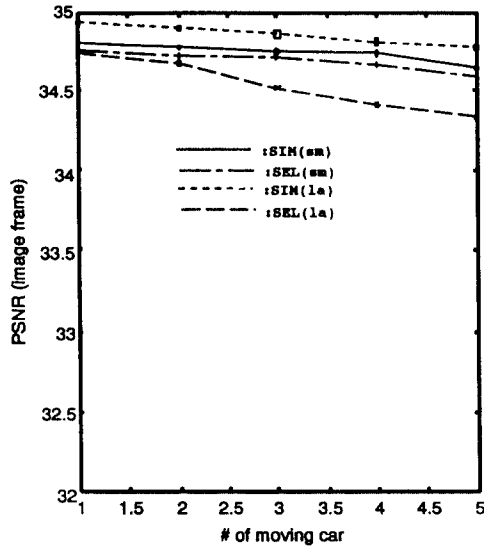


(b) large car

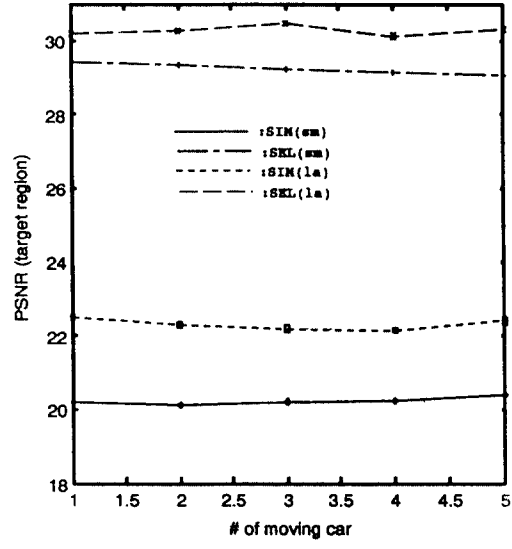


(c) target region

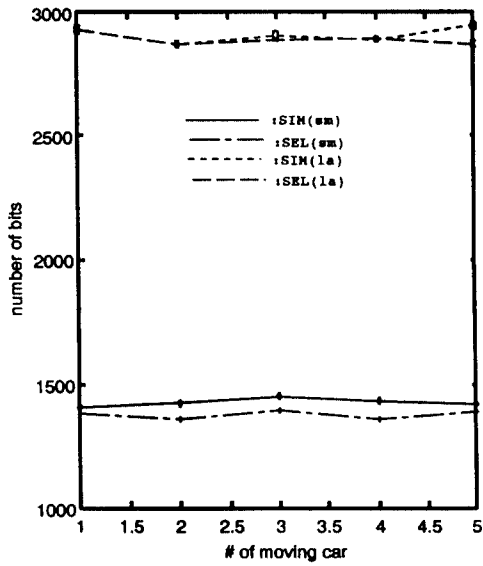
그림 7. 'Car 영상'과 목표물 영역
Fig. 7 'Car image' and target region



(a) Performance of image sequence



(b) Performance of target region



(c) transmission bits for image sequence

그림 8. 'Car 연속영상'의 PSNR 및 전송비트
 그림 8. PSNR and transmission bits for 'Car image sequence'

짐을 알 수 있다. 또, 그림 7은 'Car'의 실험 영상 및 실제 관심영역을 표시하고 있으며, 그림 8은 'Car' 영역을 인공적으로 파라미터를 변화시켜 가며 실험한 결과로서, 전체 영상의 PSNR 및 목표물 영역의 PSNR, 전송비트를 표시하고 있다. 즉 그림 8에서 SIM(sm), SIM(la)은 각각 비선택적 부호화 방식의 예로서 목표물 영역의 크기가 상대적으로 작은 경우와 큰 경우(그림 7(a)와 그림 7(b))이고, SEL(sm), SSE(la)은 각각 선택적 부호화 방식의 예로서 목표물 영역의 크기가 상대적으로 작은 경우와 큰 경우(그림 7(a)와 그림 7(b))이다.

이들 결과 역시, 제안한 방식이 다른 방식에 비하여 관심영역을 제대로 복원해내고 있음을 알 수 있다. 한편 그림 9는 관심영역의 크기가 'Table tennis' 연속영상에 비해 상대적으로 큰 경우의 일례로서, 'Trevor' 연속영상을 48kbit/s, 10Hz로 부호화했을 경우의 실험 결과를 제시하고 있다. 관심영역의 크기가 상대적으로 큰 경우에는 작은 경우에 비해서 선택적 부호화 기법의 장점이 두드러지게 드러나지는 않는다. 그러나 이경우에도 선택적 부호화를 행하지 않은 방식에 비해서 선택적 부호화 방식이 관심영역의 화질을 전반적으로 개선시키며, 연속영상 사이에서 관심영역의 화질을 보다 일정하게 유지해 줌을 알 수 있다.

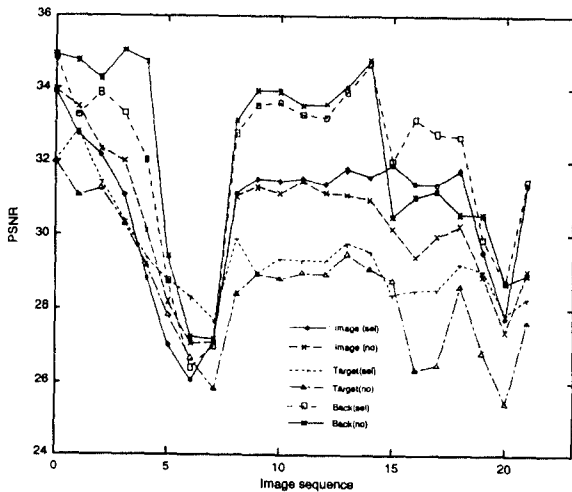


그림 9. Trevor 연속영상의 PSNR
Fig. 9 PSNR of Treor sequence

한편 관심영역이 여러 개 존재하는 경우(탁구 공과 탁구 라켓)에 대해서 살펴보자. 그림 10은 32kbit/s, 10Hz로 이들 영상을 부호화한 실험 결과를 제시하고 있는데, 이 역시 본 논문에서 제안한 '선택적 부호화 방식'이 좋은 성능을 나타내고 있다.

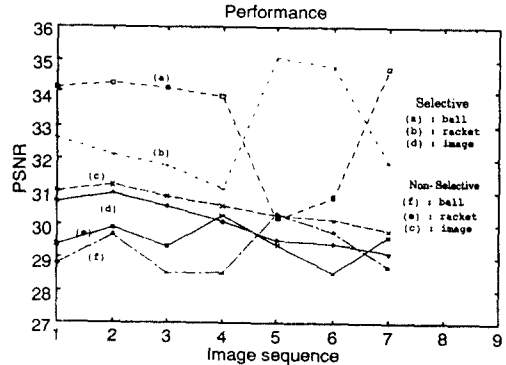


그림 10. 여러 관심영역이 존재할 경우 Table tennis 연속영상의 PSNR

Fig. 10 PSNR for Table tennis sequence where several target regions exist

표 1. PSNR 및 전송비트

Table 1. PSNR and transmission bits

	PSNR of reconstr., image (dB)	PSNR of reconstr., target (dB)	required bits for transmission
FBMA	18.21	20.64	1332
GFBMA	26.42	23.61	712
SGFBMA(I)	26.5	28.71	740
SGFBMA(II)	26.6	30.15	832

표 2. 전송비트의 성분

Table 2. Components of transmission bits

	global para.	local motion vector	quad-tree infor.	error signal (target)
FBMA		1332		
GFBMA	40	672		
SGFBMA(I)	40	62	166	472
SGFBMA(II)	40	74	190	528

만약 quadtree 기법을 적용하고 배경과 목표물을 구별하지 않고 압축을 행했을 경우는 영상 전체의 화질을 제한한 알고리즘과 비슷하거나 더 개선될 수도 있다. 그러나 우리가 중요하게 생각하는 목표물 영역에 대해서는 본 알고리즘이 훨씬 우수한 화질을 나타낸다. 그 이유는 목표물 영역에 대해서 본 알고리즘은 각종 파라미터 및 움직임 보상오차신호를 충분한 비트를 할당하여 보냄으로써 원하는 화질로 복원할 수 있도록 하기 때문이다.

감사의 글

본 논문의 일부분은 국방과학연구소(ADD) "기초연구과제"의 지원하에 이루어진 것이며, 이처럼 재정적 지원을 아끼지 않으신 국방과학연구소에 사의를 표한다.

참 고 문 헌

1. "Draft revision H.261," *CCITT SGXV*, March 1990.
2. "Simulation model for very low bit rate visual telephony(SIM3)," *CCITT SGXV, Working Party XV/1, Document \#LBC-93-025*, March 1993.
3. R. Srinivasan and K. R. Rao, "Predictive coding based on efficient motion estimation," *IEEE Trans. Commun.*, Vol. 33, 1985, pp. 888-896.
4. M. H. Hutter, "Differential estimation of the

- global motion parameters zoom and pan," *Signal Processing*, Vol. 16, 1989, pp. 249-265.
5. S.F.Wu and J.Kittler (1989), "A differential method for simultaneous estimation of rotation, change of scale and translation", *Signal Process.*, pp. 69-80.
6. D. Adolph and R. Buschmann, "1.15 Mbit/s coding of video signals including global motion compensation," *Signal Process.: Image Commun.*-3, 1991, pp. 259-274.
7. Y. Tong Tse and R.L.Baker (1991), "Global Zoom/Pan Estimation and Compensation for video Compression", *Proc. ICASSP-91*, pp. 2725-2728.
8. 이종배, 김성대, 이상지(1993), "카메라의 움직임으로 인하여 영상전체가 이동하는 경우 움직이는 목표물 추출 및 압축·전송", 1993년도 한국 통신학회 추계 종합 학술대회, 388-391.
9. 이종배, 김성대, 이상지(1993), "Quad-tree를 이용한 선택적 부호화 기법", 1994년도 한국 통신학회 추계 종합 학술대회, 1994년 7월, pp. 298-302.
10. J. B. Lee, S. D. Kim and S. Z. Lee, "Selective Coding Scheme based on motion vector information," *Proc. of EUSIPCO-94*, Sep. 1994, pp. 676-679.
11. J. B. Lee and S. D. Kim, "Moving target extraction and image coding Based on motion information", *IEICE Trans. Electronics, Communications and Computer Sciences*, Vol. E78-A, Jan. 1995, pp. 127-130.
12. R.O.Duda and P.E.Hart (1973), *Pattern Classification and Scene Analysis*, Wiley-Interscience.
13. D.H.Ballard and C.M. Brown (1982), *Computer Vision*, Prentice Hall.
14. P.Strobach, D.Schutt and W.Tengler (1988), "Space-Variant Regular Decomposition Quadrees in Adaptive Interframe Coding", *Proc. Inter. Conf. on ASSP. II*, 1988.
15. P.Strobach, (1990), "Tree-structured Scene Adaptive Coder", *IEEE Tr. on Comm.* 38.
16. G. J. Sullivan and R. L. Baker, "Efficient quad-tree coding of image and video", *IEEE Int. Conf.*

- on ASSP, May 1991, pp. 2661-2664.
17. D. J. Vaisey and A. Gersho, "Variable block size image coding", *IEEE International Conf. on ASSP*, Apr. 1987, pp.25.1.1-4.
 18. Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. on ASSP*, Vol 36, Sep. 1988, pp. 1445-1453.
 19. J.Vaisey and A. Gersho, "Image compression with variable block size segmentation", *IEEE Trans. on Signal Processing*, Vol. 40, Aug. 1992, pp. 2040-2060.

金 成 大(Seong Dae Kim)

正會員

제20권 9호 참조

현재: 한국과학 기술원 교수



李 鍾 培(Jong Bae Lee) 正會員

1962년 12월 3일생

1985년: 서울대학교 제어계측공
학과 졸업(학사)

1987년: 한국과학기술원 공학석
사 취득

1996년: 한국과학기술원 공학박
사 취득

1985년~현재: LG전자에서 근무

※주관심분야: 신호처리, MPEG-IV, internet protocol