

H.264를 위한 주파수 영역에서의 반화소 정밀도 움직임 예측 알고리즘

정회원 강민정*, 준회원 허재성*, 종신회원 류철*

Half-Pixel Accuracy Motion Estimation Algorithm in the Transform Domain for H.264

Min-Jung Kang* *Regular Member*, Jae-Seong Heo* *Associate Member*,
Chul Ryu*^o *Lifelong Member*

요약

공간 영역에서의 움직임 예측은 이전 영상에서 지정된 크기의 탐색 영역을 검색하여 현재 블록과 최소 오차를 갖는 블록을 찾는 방법으로, 탐색 영역을 검색하는 과정에서 많은 부호화 시간이 소비된다. 이러한 문제점은 공간 영역에서의 움직임 예측을 주파수 영역에서의 이동 행렬을 사용함으로써 해결할 수 있다. 본 논문에서는 기존의 이동 행렬을 새로운 재귀방정식으로 유도하여 계산량을 줄이는 동시에 영상의 화질은 기존 방식과 유사하게 유지하고자 한다. 또한 반화소 정밀도의 움직임 예측을 위하여 주파수 영역에서의 수직, 수평 이동 행렬을 간단히 수정함으로써 공간 영역에서의 양선형 보간법에 의해 더욱 증대되는 계산량 문제를 해결하고자 한다. 실험 결과 제안된 알고리즘에 의한 DCT 기반 주파수 영역에서의 움직임 예측이 공간 영역에 비하여 적은 비트량을 이용하여 보다 높은 PSNR을 제공함으로써 증명한다.

Key Words : Half-Pixel Motion Estimation, Fast Algorithm

ABSTRACT

Motion estimation and compensation in the spatial domain check the searching area of specified size in the previous frame and search block to minimize the difference with current block. When we check the searching area, it consumes the most encoding times due to increasing the complexity. We can solve this fault by means of motion estimation using shifting matrix in the transform domain instead of the spatial domain. We derive so the existed shifting matrix to a new recursion equation that we decrease more computations. We modify simply vertical shifting matrix and horizontal shifting matrix in the transform domain for motion estimation of half-pixel accuracy. So, we solve increasing computation due to bilinear interpolation in the spatial domain. Simulation results prove that motion estimation by the proposed algorithm in DCT-based transform domain provides higher PSNR using fewer bits than results in the spatial domain.

I. 서론

최근, 방송과 가전 부분에 한정되었던 영상과 음

향을 통신 및 컴퓨터와 결합하여 새로운 미디어로 탄생시킨 멀티미디어가, 멀티미디어 컴퓨터, 주문형 비디오, 고화질 TV, 디지털 방송 DVD 등 여러 가

* 동국대학교 정보통신공학과(minjong@dgu.ac.kr, neosuram@dgu.ac.kr, cryu@dgu.ac.kr)(^o: 교신저자)
논문번호 : KICS2007-10-491, 접수일자 : 2007년 10년 30일, 최종논문접수일자: 2008년 10월 6일

지 상품의 형태로 개발되고 있다. 디지털 비디오 압축 기술은 이러한 멀티미디어 응용 분야의 핵심으로, 가장 널리 사용되고 있는 표준에는 MPEG-1, MPEG-2, MPEG-4, H.261, H.263, H.264 등이 있다. MPEG-2 비디오 압축 표준은 현재 빠르게 보급되고 있는 디지털 방송 및 고화질 TV, DVD에 주로 사용되고 있고, MPEG-4는 캠코더 폰의 비디오 압축 알고리즘 및 주문형 비디오 서비스에 사용 중이다. 본 논문의 실험을 위해 사용된 H.264 비디오 압축은 핸드폰의 비디오 압축 및 주문형 비디오 서비스에 사용될 뿐만 아니라, DMB(Digital Multimedia Broadcasting)의 비디오 압축 알고리즘으로 채택되어 상용화되고 있다. 이러한 비디오 압축 표준은 일반적으로 DCT(Discrete Cosine Transform)와 움직임 예측 및 보상, 가변장 부호화의 순서로 부호화되며, 이 중 움직임 예측이 전체 동영상 부호화 시간의 60~80%를 소비함으로써 가장 많은 계산량을 차지한다¹⁾.

움직임 예측은 동영상 부호화 과정에 명시되어 있지 않기 때문에 다른 하드웨어 복잡도나 전력 소비량, 부호화 시간 등을 고려하여 다양한 알고리즘들이 개발되고 있다. 움직임 예측 알고리즘은 크게 공간 영역을 기반으로 한 방식과 주파수 영역을 기반으로 한 방식으로 구분된다. 공간 영역 알고리즘 방식은 블록 정합 알고리즘과 gradient 기반 알고리즘 등으로 구성되며²⁾⁴⁾, 주파수 영역 알고리즘은 위상 상관관계 알고리즘과 웨이블릿 변환 기반 알고리즘 그리고 DCT-기반 알고리즘으로 구성된다⁵⁾⁶⁾⁷⁾. 주파수 영역에서의 위상 상관관계 기반 움직임 예측 알고리즘은 영상 간의 정확한 움직임 예측을 제공하는 반면, 넓은 탐색 영역을 요구하기 때문에 계산상의 복잡도가 증가하며, 기존의 DCT 기반 동영상 압축 표준과의 호환성이 떨어지는 문제점을 가지고 있다. 또한 웨이블릿 변환 기반의 움직임 예측은 이웃한 데이터간의 유사성 제거 및 전체 에너지의 효율적인 이동과 압축이 가능하고, 블록킹 현상이 발생하지 않는 반면, 각 부대역에 대한 움직임 벡터값을 예측함으로써 계산량이 증대하고, 각 부대역의 움직임 벡터를 전송해야하므로 전송될 데이터량이 증가하는 단점을 갖는다. 이에 비하여, DCT-기반 움직임 예측 알고리즘은 DCT를 통하여 공간 중복성을 제거하고, 움직임 예측 및 보상을 통하여 시간 중복성을 제거하는 일반적인 DCT-기반 동영상 부호화로부터 간단히 수행될 수 있다.

일반적인 DCT 기반의 동영상 부호화 표준은 다

음의 그림 1과 같이 DCT, 양자화, 역양자화, 역 DCT, 공간 영역에서의 움직임 예측 및 보상 등으로 피드백 루프를 구성한다. 즉, DCT와 양자화, 역양자화, 역 DCT에 따라 이전 영상을 복원 후 이를 이용하여 공간 영역에서 현재 영상의 움직임을 예측한다. 반면에 주파수 영역에서의 움직임 예측은 그림 2와 같이 역 DCT 과정이 생략되고 양자화와 역 양자화 후 주파수 영역에서의 움직임 예측 및 보상이 가능하게 됨으로써, 피드백 루프는 양자화기, 역 양자화기, 그리고 주파수 영역에서의 움직임 예측 및 보상의 함수들로만 구성됨에 따라 부호화 시간을 절약할 수 있다. 또한 영상의 움직임 예측이 정수 단위에서 반화소 단위로 확장될 경우, 공간 영역에서는 영상을 가로와 세로 방향으로 각각 2배 확대하여 움직임을 예측하기 때문에 정수 단위에 비해 상대적으로 계산량이 상당히 늘어나게 된다. 본 논문에서는 공간 영역에서의 반화소 단위 움직임 예측 과정에서 발생하는 계산상 복잡도 증가 문제를 주파수 영역에서 해결하여 동영상 부호화 시간을 단축시키고자 새로운 방식을 제안한다. 본 논문의 구성은 제 II장에서 DCT-기반의 움직임 예측 알고리즘을 설명하고, 제 III장은 본 논문에서 제안하는 재귀방정식을 유도한다. 제 IV장에서는 실험

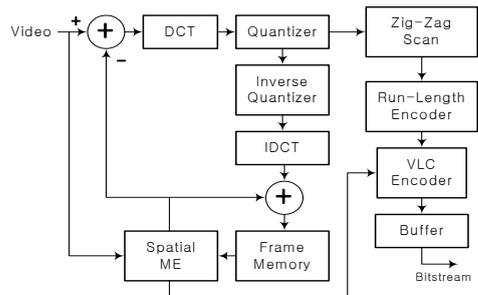


그림 1. 공간 영역에서의 움직임 예측

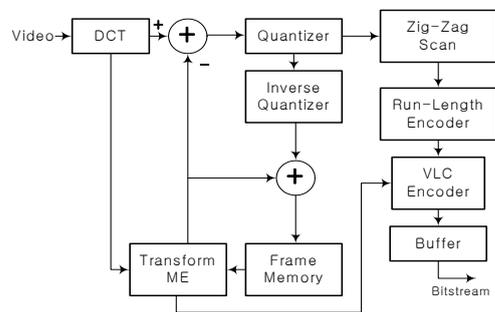


그림 2. 주파수 영역에서의 움직임 예측

결과를 통하여 제안한 재귀방정식이 공간 영역에서의 움직임 예측 방식보다 PSNR과 발생 데이터 측면에서 더 좋은 결과를 나타냄을 보여준다. 마지막으로 제 V장에서는 본 논문에서 제안된 방법을 간단히 정리하고 끝을 맺는다.

II. 주파수 영역에서의 움직임 예측

변환 부호화는 압축하려는 정보의 내용을 잃지 않으면서 압축 효율을 높이기 위하여 정보의 표현 방법을 바꾸는 기술로써, 영상 압축으로 가장 널리 사용되고 있는 방법이 DCT 변환이다. 공간 영역에서의 각 화소들은 그 주변 화소값들과 서로 비슷하기 때문에 화소간의 상관관계를 갖는다. DCT 변환은 원영상을 저주파와 고주파의 주파수 성분으로 분해하여, 각 화소간의 상관관계를 없애고, 입력 영상의 에너지가 저주파대에 집중되는 특성을 갖는다. 또한 인간 눈의 시각적 특성은 저주파 성분에 민감하고 고주파 성분에는 둔감하여 고주파 성분의 정보를 제거하더라도 저주파 성분의 변화보다 잘 인지하지 못하기 때문에 DCT의 이러한 특성은 영상 압축을 용이하게 한다. 푸리에 분석에 따르면, 어떠한 경우의 신호라도 다양한 크기와 주파수를 갖는 코사인 파들의 합으로 표현될 수 있고, 이를 1차원으로 다음의 식 (1)과 (2)로 나타낼 수 있다^[8].

$$y(n) = \sum_{k=0}^{N-1} t(k,n)x(n), \quad 0 \leq k \leq N-1 \quad (1)$$

$$t(k,n) = \begin{cases} \frac{1}{\sqrt{N}}, & k=0 \\ \sqrt{\frac{2}{N}} \cos \frac{\pi(2n+1)k}{2N}, & 1 \leq k \leq N-1 \end{cases}, \quad 0 \leq n \leq N-1 \quad (2)$$

일반적인 DCT 변환을 이용한 정지 영상과 동영상 압축 표준들은 8x8 크기의 DCT 블록을 사용하는 반면, H.264는 4x4 직교 변환을 사용한다. 4x4 화소 단위의 직교 변환은 8x8 화소 단위보다 계산되는 데이터수가 적고, 연산의 유효 자리수가 적기 때문에 구현이 용이하다. 또한 H.264 움직임 예측에서의 부호화 최소 단위가 4x4 화소이므로 직교 변환 역시 4x4 화소 단위로 수행된다. 따라서 본 논문에서도 4x4 화소 단위의 직교 변환을 이용하여 반화소 단위 움직임 예측을 수행하고자 N이 4인 경우를 고려한다. 기존의 정수 단위 움직임 예측은 Plompen이 제안한 이동 행렬 알고리즘을 이용하여 이전 블록의 행렬과 이동 행렬의 곱으로 수행될 수

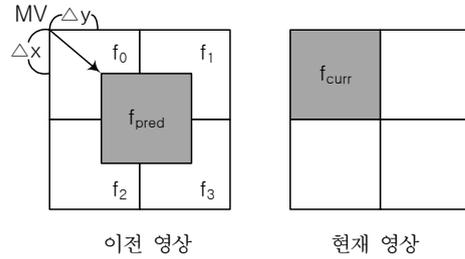


그림 3. 움직임 예측

있다^[9]. Plompen이 제안한 이동 행렬 기반 알고리즘은 이전의 참조 영상에서 예측한 f_{pred} 는 주변 4개 블록인 f_0, f_1, f_2, f_3 의 수평, 수직 변환된 형태의 합으로써 표현 가능함을 보여주었다. 그림 3은 현재 블록 f_{curr} 에 대하여 가장 정확히 예측된 f_{pred} 와 그에 따른 움직임 벡터의 관계를 보여준다.

예측된 f_{pred} 는 식 (3)에서 보는 바와 같이 수직 이동 행렬 V_i 와 수평 이동 행렬 H_i 를 이용하여 계산되며, 이때의 수직 이동 행렬 V_i 와 수평 이동 행렬 H_i 는 각각 식 (4)와 (5)이다.

$$f_{pred} = \sum_{i=0}^3 V_i f_i H_i \quad (3)$$

$$V_0 = D_{4-\Delta x}, V_1 = D_{4-\Delta x}, \quad (4)$$

$$V_2 = D_{\Delta x}^t, V_3 = D_{\Delta x}^t \quad (5)$$

$$H_0 = D_{4-\Delta y}^t, H_1 = D_{\Delta y},$$

$$H_2 = D_{4-\Delta y}^t, H_3 = D_{\Delta y}$$

식 (4)와 (5)에 사용된 변위 행렬 D_n 은

$$D_n = \begin{bmatrix} | & & & \\ 0 & I_n & & \\ | & & & \\ - & - & - & - \\ 0 & 0 & & \end{bmatrix} \quad (6)$$

로 정의되며, I_n 은 $n \times n$ 단위행렬이다.

또한 DCT 변환은 식 (8)과 식 (9)에서 보는 바와 같이 직교성과 분리성을 갖는다.

$$DCT(A) \cong \hat{A} = T A T^t \quad (7)$$

$$T \cdot T^t = I \quad (8)$$

$$DCT(ab) = DCT(a)DCT(b) = \hat{a}\hat{b} \quad (9)$$

따라서 이러한 DCT 변환의 특성을 이용하여 식 (3)의 예측된 블록 f_{pred} 의 DCT 계수들을 다음과 같이 전개할 수 있다.

$$\begin{aligned} \widehat{f}_{pred} &= DCT(f_{pred}) = DCT\left(\sum_{i=0}^3 V_i f_i H_i\right) \\ &= \sum_{i=0}^3 DCT(V_i) DCT(f_i) DCT(H_i) \\ &= \sum_{i=0}^3 (TV_i T^t)(Tf_i T^t)(TH_i T^t) \\ &= \sum_{i=0}^3 \widehat{V}_i \widehat{f}_i \widehat{H}_i \end{aligned} \quad (10)$$

$$\widehat{V}_i = TV_i T^t, \widehat{f}_i = Tf_i T^t, \widehat{H}_i = TH_i T^t \quad (11)$$

식 (10)은 예측된 블록 f_{pred} 의 DCT 계수들을 이전의 주변 4개 블록 f_0, f_1, f_2, f_3 의 DCT 계수들과 미리 계산된 수직 이동 행렬 V_i 와 수평 이동 행렬 H_i 의 DCT 계수들을 이용하여 구할 수 있음을 보여준다. 주파수 영역에서의 움직임 예측 시, 현재의 DCT 계수 블록 f_{curr} 과 가장 정확히 예측된 DCT 계수 블록 f_{pred} 의 상대적 위치로부터 움직임 벡터를 구하고, f_{curr} 와 f_{pred} 간의 차이값은 다시 부호화 과정을 거치게 된다.

III. 반화소 움직임 예측을 위한 주파수영역 재귀방정식 유도

H.264 표준에서는 여러 장의 참조 영상을 이용하여 7가지 모드에 대하여 움직임 예측을 수행함으로써 기존의 동영상 압축 부호화에 비해 더 좋은 압축 효율을 제공한다. 만약 입력 영상의 움직임이 큰 경우에는 여러 장의 참조 영상을 사용할 때 보다 좋은 움직임 예측이 가능한 반면, 참조 영상의 수에 비례하여 연산량이 증가하게 된다. 본 논문에서는 이러한 연산량 증대 문제를 해결하기 위하여 여러 장의 참조 영상 중 각 모드별 움직임 예측에 사용될 참조 영상을 선택하는 방법을 제안한다. 제안하는 방식은 16×16 모드에서 선택된 최적의 참조 영상이 16×16 모드 이하의 다른 모드에서도 참조 영상이 될 가능성이 높다는 점과 화면 간 예측 모드의 상관성을 고려하여 모드 선택 시 방향성이 유지된다는 점을 이용한다. 즉, 참조 영상의 개수를 제한하는 방식과 화면 간 예측 모드의 개수를 제한하는 방식을 동시에 적용하여 기존 알고리즘에 비해 더욱 연산량을 줄이고자 한다.

주파수 영역에서 사용되는 움직임 예측 알고리즘의 가장 큰 문제는 높은 연산량에 있다. 이전 영상의 주변 블록 f_0, f_1, f_2, f_3 의 DCT 계수들을 이용

하여 예측된 블록 f_{pred} 의 DCT 계수들을 구하기 위해서는 8개의 행렬 곱셈과 3개의 행렬 덧셈이 요구된다. 그러나 DCT 변환 특성인 주파수 영역에서 저주파 영역으로 에너지가 집중되는 특성과 DCT 변환 후 양자화된 블록 계수들이 대부분 0이 된다는 희소성을 이용하면 연산량을 효과적으로 줄일 수 있다.

$\widehat{f}_i(m,n)$ 은 블록 \widehat{f}_i 의 m 번째 행과 n 번째 열의 행렬 성분이며, 열벡터 $\widehat{V}_i(m)$ 은 행렬 V_i 의 m 번째 열을 그리고, 행벡터 $\widehat{H}_i(n)$ 은 행렬 H_i 의 n 번째 행을 나타낸다. DCT 변환 후 양자화된 블록 \widehat{f}_i 의 특성을 기반을 두어 계산상의 복잡도를 줄이기 위해 식 (10)을 다음과 같이 유도할 수 있다.

$$\begin{aligned} \widehat{f}_{pred} &= \sum_{i=0}^3 \widehat{V}_i \widehat{f}_i \widehat{H}_i \\ &= \sum_{i=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 \widehat{f}_i(m,n) \widehat{V}_i(m) \widehat{H}_i(n) \\ &= \sum_{m=0}^3 \sum_{n=0}^3 \sum_{i=0}^3 \widehat{f}_i(m,n) \widehat{V}_i(m) \widehat{H}_i(n) \end{aligned} \quad (12)$$

$$\begin{aligned} \widehat{f}_{pred} &= \sum_{m=0}^3 \sum_{n=0}^3 \sum_{i=0}^3 \widehat{f}_i(m,n) \widehat{V}_i(m) \widehat{H}_i(n) \\ &= \widehat{f}_0(0,0) \widehat{V}_0(0) \widehat{H}_0(0) + \widehat{f}_1(0,0) \widehat{V}_1(0) \widehat{H}_1(0) \\ &\quad + \widehat{f}_2(0,0) \widehat{V}_2(0) \widehat{H}_2(0) + \widehat{f}_3(0,0) \widehat{V}_3(0) \widehat{H}_3(0) \\ &\quad + \widehat{f}_0(0,1) \widehat{V}_0(0) \widehat{H}_0(1) + \widehat{f}_1(0,1) \widehat{V}_1(0) \widehat{H}_1(1) \\ &\quad + \dots \\ &\quad + \widehat{f}_0(3,3) \widehat{V}_0(3) \widehat{H}_0(3) + \widehat{f}_1(3,3) \widehat{V}_1(3) \widehat{H}_1(3) \\ &\quad + \widehat{f}_2(3,3) \widehat{V}_2(3) \widehat{H}_2(3) + \widehat{f}_3(3,3) \widehat{V}_3(3) \widehat{H}_3(3) \end{aligned} \quad (13)$$

식 (12)에서 유도된 \widehat{f}_{pred} 는 $\widehat{V}_i(m)$ 과 $\widehat{H}_i(n)$ 의 외적에 의하여 얻어진 베이스 영상의 선형 조합으로 간주될 있고, $\widehat{f}_i(m,n)$ 이 0값을 갖는 외적의 경우에는 식 (13)의 계산에서 제외될 수 있기 때문에 \widehat{f}_i 의 희소성을 이용하여 \widehat{f}_{pred} 을 효과적으로 계산할 수 있다. 이는 일반적으로 저주파 대역에 있는 몇몇 $\widehat{f}_i(m,n)$ 만이 0이 아닌 값을 갖고, 고주파 대역에 있는 $\widehat{f}_i(m,n)$ 의 대부분이 0의 값을 갖기 때문이다. 따라서 식 (12)는 지그재그 스캔 순서에 따라 식 (13)과 같이 정리할 수 있다. 식 (13)에서 0이 아닌 $\widehat{f}_i(m,n)$ 만이 $\widehat{V}_i(m)$ 과 $\widehat{H}_i(n)$ 의 외적을 스케일링하게 되고, \widehat{f}_{pred} 의 최종 결과를 얻기 위하여 다음과 같이 나타낼 수 있다.

$$\widehat{f} = \widehat{f}_{pred}^{k-1} + \widehat{f}_i(m,n) \widehat{V}_i(m) \widehat{H}_i(n) \quad (14)$$

여기서 \hat{f}_{pred}^0 는 4x4의 영행렬이며, k 는 1부터 f_0, f_1, f_2, f_3 의 0이 아닌 DCT 계수들의 전체 개수까지 증가한다.

식 (14)에서 제안된 재귀방정식은 두 가지 중요한 점을 가진다. 첫째, 저주파 성분으로 구성된 영상부터 저주파와 고주파 성분으로 구성된 영상 등으로 다른 주파수 대역을 갖는 움직임 예측된 영상은 $\hat{f}_i(m,n)$ 의 개수를 조절하여 효율적으로 생성 가능하다. 둘째, $\hat{V}_i, \hat{H}_i, \hat{f}_i$ 의 $N \times N$ 크기가 증가하고, \hat{f}_i 의 0이 아닌 성분의 총 개수 m 이 감소할수록, 식 (14)의 재귀방정식은 연산량을 효율적으로 줄일 수 있다. 식 (10)과 본 논문에서 제안한 식 (13)의 연산량을 계산해보면 다음과 같다. 식 (10)의 일반적인 행렬 연산에서 전체 곱셈량은 $8N^3$ 이고, 덧셈량은 $(8N-5)N^2$ 이 된다. 그러나 식 (13)에서는 \hat{f}_i 의 각 0이 아닌 성분의 총 곱셈량은 $\hat{f}_i(m,n)$ 와 $\hat{V}_i(m)$ 의 곱셈량이 N 이고, $\hat{f}_i(m,n)\hat{V}_i(m)$ 과 $\hat{H}_i(n)$ 의 곱셈량이 N^2 이 되어 N^2+N 이 된다. 따라서 $0 \leq i \leq 3$ 이고, 각 \hat{f}_i 의 0이 아닌 성분의 개수가 m 인 경우, 식 (13)에 필요한 전체 곱셈량과 덧셈량은 각각 $4mN^2+4mN$ 과 $(4m-1)N^2$ 이 된다. 그러므로 식 (14)의 재귀방정식은 일반적인 방법에 비해 DCT 변환 후 양자화된 블록 \hat{f}_i 의 희소성을 충분히 이용함으로써 효율적으로 연산량을 줄일 수 있게 된다. 매크로블록별 공간 영역과 주파수 영역에서의 움직임 추정에 사용되는 연산량의 비교를 표 1을 통해 알 수 있다. 공간 영역의 경우 추정을 위해 모드별로 각각의 SAD를 구해야 하지만 주파수 영역의 경우 모드와 상관없이 기본적으로 4×4 의 크기별로 추정을 수행하므로 처음 16×16 모드를 위해 움직임 추정 시 얻은 SAD 값을 재사용할 수 있다¹⁰⁾.

또한 식 (14)를 이용하여 주파수 영역에서 반화소 단위 움직임 예측 또한 가능하다. 이전 영상과 현재 영상간의 시간적 중복성을 보다 효과적으로 제거하기 위해 정수 화소가 아닌 반화소 정밀도를 갖는 움직임 예측 기법들이 많이 연구되어 왔다. H.263과 MPEG-4, H.264와 같은 동영상 부호화 표

준은 반화소 정밀도를 이용한 움직임 벡터를 제공하고 있으며, 이러한 연산은 더욱 높은 계산의 복잡도와 메모리 크기의 증가를 요구한다. 일반적인 반화소 움직임 예측은 가장 인접한 정수 위치의 화소 값들을 이용하여 양선형 보간법으로 2배 확대된 영상에 대하여 움직임 예측을 수행한다. 즉, 가로 방향과 세로 방향으로 각각 2배씩 확대된 영상을 이용하여 움직임 예측을 수행하기 때문에 정수 화소 단위 움직임 예측과 비교할 때 더욱 많은 연산량을 필요로 하게 된다. 그러나 본 논문에서 유도한 식 (15)~(18)의 일반화된 수직 이동 행렬 V_i 와 수평 이동 행렬 H_i 를 사용할 경우, 추가비용 없이 주파수 영역에서의 반화소 정밀도 움직임 예측이 가능하다.

$$V_0 = V_1 = \frac{1}{2}(D_{4-\lfloor \Delta x \rfloor} + D_{4-\lceil \Delta x \rceil}) \quad (15)$$

$$V_2 = V_3 = \frac{1}{2}(D_{\lfloor \Delta x \rfloor}^t + D_{\lceil \Delta x \rceil}^t) \quad (16)$$

$$H_0 = H_2 = \frac{1}{2}(D_{4-\lfloor \Delta y \rfloor}^t + D_{4-\lceil \Delta y \rceil}^t) \quad (17)$$

$$H_1 = H_3 = \frac{1}{2}(D_{\lfloor \Delta y \rfloor} + D_{\lceil \Delta y \rceil}) \quad (18)$$

$\lfloor x \rfloor$ 는 버림함수를, $\lceil x \rceil$ 는 올림함수를 의미한다.

IV. 실험 결과

주파수 영역에서의 움직임 예측 알고리즘의 실험을 위하여 H.264 동영상 부호화기를 사용하였다. DCT는 공간 영역의 신호를 주파수 영역으로 변환하여 에너지를 저주파대로 집중시킨다. 따라서 양자화 후 대부분의 고주파 성분들은 0이 되고, 저주파 성분의 몇몇 DCT 계수들만이 존재하게 된다. 양자화 파라미터가 커짐에 따라 대부분의 고주파 대역의 DCT 계수들은 0이 되며 0이 아닌 DCT 계수의 수는 상당히 감소하게 된다. 이러한 특성을 재귀방정식을 이용한 움직임 예측에 적용함으로써 연산량을 효과적으로 줄일 수 있다. 그림 4는 carphone QCIF 영상의 두 번째 영상을 양자화 파라미터 45로 설정하여 실험하였을 때, 제안된 재귀방정식을 이용한 각 블록별 곱셈량과 덧셈량을 보여준다. 식 (10)에 의한 행렬 연산의 경우, 블록 크기가 4 ($N=4$)일 때 곱셈량과 덧셈량은 각각 512회와 432회로 모든 블록에서 동일하다. 그러나 제안한 재귀방정식을 이용한 경우, 지그재그 스캔 순서에 따라 0이 아닌 계수만이 연산에 포함됨으로써, 각 블록별

표 1. 매크로블록별 연산량 비교(p: 탐색구간, N: 블록크기)

모드선택	16x16, 16x8, 8x16	8x8, 8x4, 4x8, 4x4
공간 영역	$16p^2N^2 + 2 \times 8p^2N^2 + 4 \times 4p^2N^2$	$4p^2N^2 + 2 \times 2p^2N^2 + 4p^2N^2$
주파수 영역	$16p^2N^2 + 2 \times 8p^2 + 4 \times 4p^2$	$4p^2 + (2 \times 2p^2) + 4p^2$

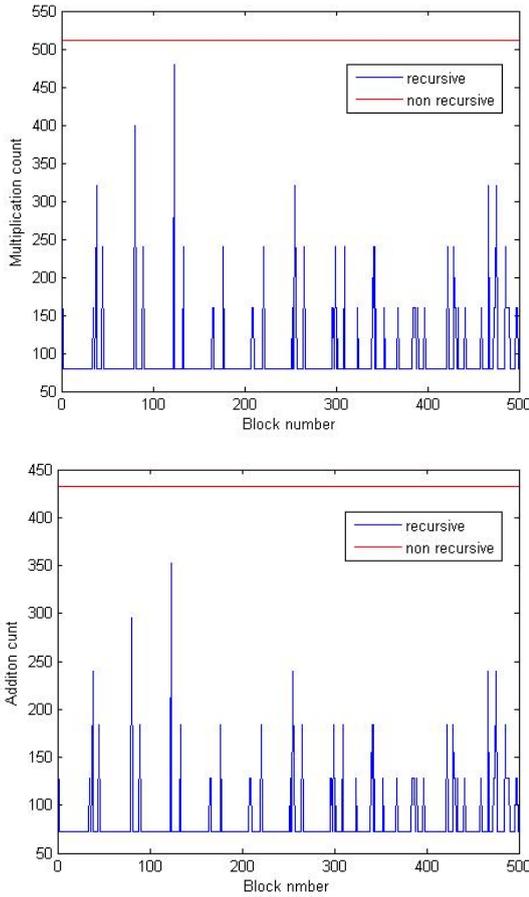


그림 4. 제안된 재귀방정식을 이용한 각 블록별 곱셈, 덧셈 연산량(Carphone QCIF의 2번째 영상, 양자화 파라미터=45)

0이 아닌 계수의 수(m)가 변동되면서 연산량 또한 많은 차이를 갖게 된다. 즉, 표 2에서 한 장의 프레임에 대한 0이 아닌 계수의 평균 개수는 6개이고, 평균 곱셈량과 덧셈량은 각각 442회와 338회로, 식

(10)에 비해 곱셈량은 14%, 덧셈량은 22% 감소하였다.

본 실험에서는 공간 영역과 DCT 기반 주파수 영역에서의 전역 탐색 기법을 이용하여 움직임 예측을 수행하였으며, 알고리즘의 성능 비교는 H.264를 위한 주파수 영역에서의 움직임 예측과 관련된 기존 논문이 없는 관계로, 정수 화소 단위와 반화소 단위 움직임 예측의 결과만을 비교하였다. 또한 H.264에서 사용되는 직교 변환의 단위가 4×4 서브블록이므로, 본 실험에서는 H.264의 가변 블록 움직임 예측을 적용하지 않고, 4×4 서브블록에 대한 움직임 예측만을 수행하였다. 이때 움직임 벡터를 결정하기 위한 기준으로는 다음의 식 (19)와 (20)을 이용하였다. 여기서 SAD_{sp} 와 SAD_{tr} 은 각각 공간 영역과 주파수 영역에서 사용된 기준 조건이다.

$$SAD(\Delta x, \Delta y)_{sp} = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} |f_{curr}(u, v) - f_{pred}(u, v)| \tag{19}$$

$$SAD(\Delta x, \Delta y)_{tr} = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} |\widehat{f}_{curr}(u, v) - \widehat{f}_{pred}(u, v)| \tag{20}$$

표 2과 표 3는 QCIF 형식의 foreman과 claire 영상에 대하여 양자화 파라미터를 변화시켜 각 공간 영역과 주파수 영역에서의 정수 화소 단위와 반화소 단위 움직임 예측 결과를 나타낸다. foreman 영상에서 주파수 영역의 정수 단위 가변블록 움직임 예측을 수행한 경우, 공간 영역과 비교하여 비트량이 최소 91 bits에서 최대 1,678 bits까지 감소한 반면, 영상의 화질을 표현하는 PSNR은 평균 0.152dB 증가하였다. 이는 주파수 영역의 움직임

표 2. 공간 영역과 주파수 영역에서의 정수화소 단위 움직임 예측 비교 (단위: PSNR-dB, 비트량-bits)

양자화 파라미터			3	6	9	12	15	18	21	24	27	30
공간 영역	Foreman	PSNR	61.51	55.80	52.79	50.28	47.88	45.24	42.88	40.41	38.48	36.52
		비트량	93,392	83,129	72,609	62,451	50,154	38,901	29,248	20,541	14,160	9,582
	Claire	PSNR	61.48	56.34	53.56	51.65	50.41	48.36	46.24	44.00	41.80	39.88
		비트량	60,251	49,822	35,176	23,756	14,692	11,444	8,571	6,497	4,961	3,682
주파수 영역	Foreman	PSNR	61.72	55.89	52.93	50.42	47.98	45.35	43.01	40.65	38.67	36.69
		비트량	93,129	81,881	71,728	60,773	49,115	38,395	28,405	19,948	13,965	9,491
	Claire	PSNR	61.80	56.61	53.90	52.00	50.47	48.56	46.36	44.27	41.92	39.89
		비트량	60,610	46,667	34,246	23,266	15,097	10,541	7,634	5,714	4,152	3,602

표 3. 공간 영역과 주파수 영역에서의 반화소 단위 움직임 예측 비교 (단위: PSNR-dB, 비트량-bits)

양자화 파라미터			3	6	9	12	15	18	21	24	27	30
공간 영역	Foreman	PSNR	61.33	55.91	52.34	50.01	48.17	45.38	43.04	40.43	38.42	36.48
		비트량	89,130	78,778	68,295	57,713	42,204	34,256	24,701	17,053	11,259	7,356
	Claire	PSNR	61.50	56.41	53.65	51.69	50.48	48.63	46.56	44.57	42.32	40.53
		비트량	58,683	48,383	33,872	22,804	13,718	10,397	7,495	5,489	4,138	3,065
주파수 영역	Foreman	PSNR	61.54	55.90	52.57	50.19	48.41	45.72	43.02	40.97	38.84	37.10
		비트량	87,025	75,942	55,811	53,212	41,992	31,023	22,468	15,263	9,989	6,487
	Claire	PSNR	61.79	56.12	54.02	52.31	50.69	48.32	46.61	45.18	42.62	40.22
		비트량	54,084	40,001	27,608	19,181	12,486	9,154	6,849	5,041	3,861	2,921

예측의 경우, 동영상상을 부호화할 때 사용된 데이터 양은 공간 영역에 비해 적게 사용된 반면, 영상의 화질은 공간 영역과 유사하게 유지됨을 의미한다. 표 3의 claire 영상에 대한 반화소 단위 움직임 예측 114 bits에서 최대 4,599 bits 감소한 반면, PSNR은 최소 0.05dB에서 최대 0.61dB까지 증가한 것을 확인할 수 있다. 그림 5와 그림 6은 각각 양자화 파라미터를 15로 고정하고 총 50장의 QCIF 형식의 carphone 영상을 이용한 비트량과 PSNR을 보여준다. 그림 5에서 가장 아래쪽 그래프부터 주파수 영역에서의 반화소 단위 움직임 예측, 공간 영역에서의 정수화소 단위 움직임 예측 그리고 공간 영역에서의 정수화소 단위 결과이다. 즉, 정수화소 단위보다는 반화소 단위 움직임 예측에 적은 비트가 사용되었고, 같은 화소 단위의 경우 본 논문에서 제안한 재귀 방정식을 이용한 주파수 영역에서의 움직임 예측이 공간 영역에서보다 적은 양의 비트가 사용

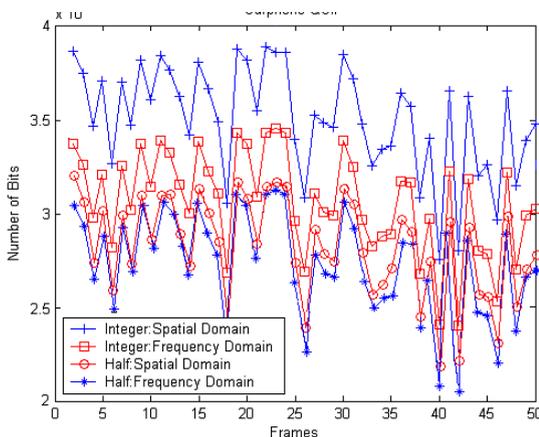


그림 5. 50장의 각 영상에 대한 비트량(양자화 파라미터=15)

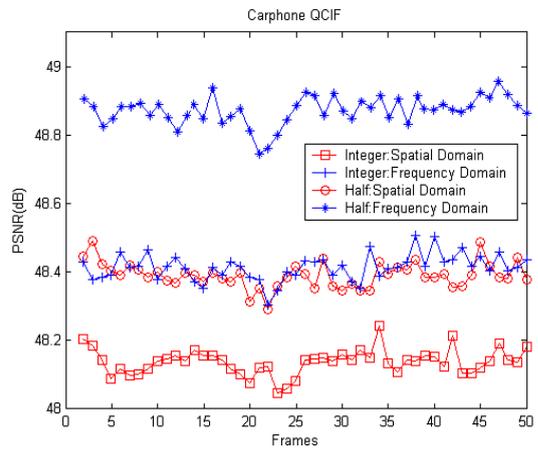


그림 6. 50장 각 영상에 대한 PSNR(양자화 파라미터=15)

되었음을 확인할 수 있다. 또한 그림 6에서 보는 바와 같이 같은 화소 단위에서 공간 영역과 주파수 영역에 대한 PSNR을 비교해볼 때, 본 논문에서 제안한 방식을 이용하여 주파수 영역에서 움직임 예측을 수행한 결과 PSNR이 더 높게 나타남을 확인할 수 있다.

V. 결론

본 논문에서는 주파수 영역에서의 반화소 단위 움직임 예측을 보다 효과적으로 수행하기 위한 새로운 알고리즘과 연산 구조를 제안하였다. 제안된 움직임 예측 알고리즘은 재귀방정식을 이용하여 계산상의 복잡도를 크게 감소시켜 계산량을 줄일 수 있었다. 또한 기존 공간 영역에서의 움직임 예측에 비해 본 논문에서 제안한 방식을 이용할 경우 부호화에 사용된 비트량이 줄어들더라도 그에 따른 화

질은 저하되지 않음을 확인할 수 있었다. 다양한 특성을 갖는 동영상에 이용한 모의실험을 통하여 주파수 영역에서의 움직임 예측 방식은 공간 영역에서의 움직임 예측보다 높은 PSNR과 압축률을 보여 주었다.

참 고 문 헌

[1] P.Kuhn, "Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation," Boston, MA:Kluwer Academic, 1999.

[2] J. Zan, M. O. Ahmad and M. N. S. Swamy, "Pyramidal motion estimation techniques exploiting Intra-level motion correlation," IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing, Vol.50, No.2, pp.83-93, Feb. 2003.

[3] Y. Keller and A. Averbuch, "Fast gradient methods based on global motion estimation for video compression," IEEE Trans. Circuits and Systems for Video Technology, Vol.13, No.4, pp.300-309, Apr. 2003.

[4] C. Du, Y. He and J. Zheng, "PPHPS: A parabolic prediction-based, fast half-pixel search algorithm for very low bit-rate movint-picture coding," IEEE Trans. Circuits and Systems for Video Technology, Vol.13, No.6, pp.514-518, Jun. 2003.

[5] J. Zan, M. O. Ahmad and M. N. S. Swamy, "Comparison of wavelets for multiresolution motion estimation," IEEE Trans. Circuits and Systems for Video Technology, Vol.16, No.3, pp.439-446, Mar. 2006.

[6] N. Merhav and V. Bhaskaran, "A fast algorithm for DCT-domain inverse motion compensation," IEEE International Conference on Acoustics, Speech and Signal Processing, Vol.4, pp.2307-2310, May 1996.

[7] J. Song and B. Yeo, "A fast algorithm for DCT-domain inverse motion compensation based on shared information in a macroblock," IEEE Trans. on Circuits and Systems for Video technology, Vol.10, No.5, pp.767-775, Aug. 2000.

[8] K. R. Rao and P. Yip, Discrete Cosine Transform, Academic Press, 1990.

[9] R. H. J. M. Plompen, B.F.Schuurink and J.Biemond, "A new motion-compensated transform coding scheme," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Vol.1, pp.371-374, 1985.

[10] Yun Wu, Yong Zhao, Jianshi Li, "Brief Analysis of the H.264 Coding Standard", Intelligent Information Hiding and Multimedia Signal Processing, 2007. IHHMSP 2007. Third International Conference on Volume 2, 26-28 pp.154-157, Nov. 2007.

강 민 정 (Min-Jung Kang)

정회원



2001년 2월 동국대학교 정보통신공학과 (공학사)
 2003년 2월 동국대학교 정보통신공학과 (공학석사)
 2008년 현재 동국대학교 정보통신공학과 박사 수료
 <관심분야> 데이터 압축, 오류 제어, 멀티미디어 영상 통신

허 재 성 (Jae-Seong Heo)

준회원



2007년 2월 동국대학교 정보통신공학과 (공학사)
 2008년 현재 동국대학교 정보통신공학과 석사 수료
 <관심분야> 데이터 압축, 멀티미디어 영상통신

류 철 (Chul Ryu)

종신회원



1989년 Florida Institute of Technology 컴퓨터공학과 공학사
 1991년 Polytechnic University 전기공학과 공학석사
 1997년 Polytechnic University 전기공학과 공학박사
 1998년~1999년 LG 정보통신연구소 선임연구원
 1999년~현재 동국대학교 IT학부 부교수
 <관심분야> Visual Communication, High Resolution Reconstruction, Joint Source Coding