

웨이블렛 변환 영역에서 다중 해상도 움직임 추정에 관한 연구

이진환^{*,**}, 김소연^{*}, 노용만^{*}

^{*}한국정보통신대학원, 멀티미디어그룹

^{**}한국전자통신연구원, 방송기술연구부

A Study on Multiresolution Motion Estimation on the Wavelet Transform Domain

Jinhwan Lee^{*,**}, Soyeun Kim and Yong Man Ro^{*}

^{*} Multimedia Group/Information and Communication University (ICU)

^{**} Broadcasting Technology Department/Electronics and Telecommunications Research Institute (ETRI)

Abstract

Wavelet transform decomposes a video frame into a set of subframes with different resolutions corresponding to different frequency bands. The block size of highest layer subframe is too small to estimate correct motion. To decrease this aperture problem in the approximation frame, multiresolution motion estimation (MRME) using hierarchical block matching is presented in this paper. In the multiresolution motion estimation approach, motion vectors in the subframes are estimated by the motion vectors in the approximation frame and are refined at each subframe. It reduces computational time in motion estimation. It is shown that our MRME approach has a superior performance than traditional MRME approaches.

$$f(x,y) \xrightarrow{W} \{ A_f, H_f, V_f, D_f, H_m, f, \dots, H_f, V_f, D_f \} \quad (1)$$

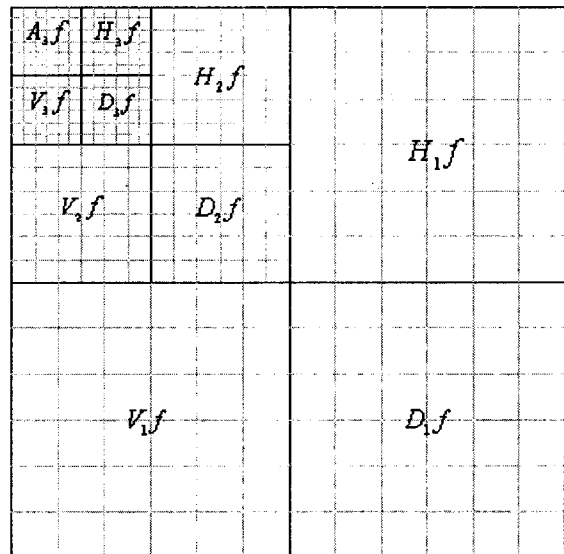
여기서 $f(x,y)$ 는 원 영상을 의미하며, W 는 2-D DWT를 뜻한다. $m(m>0)$ 은 웨이블렛 변환에서 해상도의 깊이를 의미하는 것으로 식 (1)은 m 계층 웨이블렛 변환이라 일컬으며, 2^{-1} 에서 2^{-m} 까지의 해상도를 갖는 영상들로 구성된다. 이때 A_f 는 기초(approximation) 영상을, H_f, V_f, D_f 는 각각 수평(horizontal), 수직(vertical), 대각(diagonal) 성분으로 세밀한 정보를 지니고 있는 영상을 의미한다. 즉, 웨이블렛 변환은 <그림 1>과 같이 원 영상을 다중 해상도와 주파수를 갖는 여러장의 서브영상으로 분해한다.

I. 웨이블렛 부호화 및 움직임 추정

1) 웨이블렛 부호화

웨이블렛 변환은 영상 신호를 공간과 주파수 영역에서 다중 해상도 및 다중 주파수로 표현할 수 있는 특징을 지니고 있다. 웨이블렛 변환이 갖는 이러한 계층적 구조 특성은 원 신호를 다양한 해상도와 비트율로 부호화하여 전송할 수 있다. 따라서 웨이블렛으로 부호화된 영상신호들은 QCIF(Quarter Common Intermediate Format), CIF, SDTV(Standard Definition Television), HDTV(High Definition Television)와 같은 여러 종류의 영상 포맷 사이에서 유동적으로 사용되어질 수 있다. 또한 웨이블렛 방식의 부호화는 DCT(Discrete Cosine Transform)방식에 의한 영상 부호화에서 흔히 야기될 수 있는 블록킹 현상을 줄일 수 있는 장점을 지닌다. 따라서 최근에 멀티미디어 국제 표준화에서는 정지영상 압축 표준이 될 JPEG-2000(Joint Photographic Experts Group-2000)의 기본 부호화 방식으로 웨이블렛을 채택한 바 있다.

영상에 대한 2-D DWT(2-Dimensional Discrete Wavelet Transform)는 다음 식과 같이 표현될 수 있다.

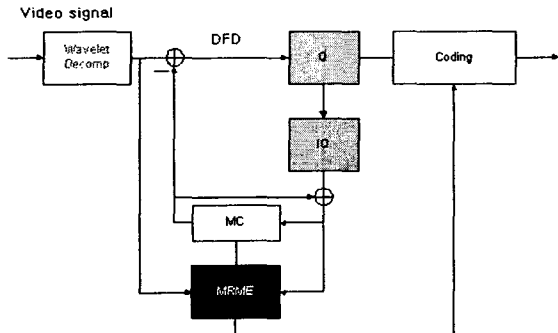


<그림 1> 웨이블렛 변환(3 계층 DWT)

2) 웨이블렛을 이용한 영상부호화기

<그림 2>는 웨이블렛 변환을 이용한 영상 부호화기의 블록 다이어그램이다. 입력된 영상신호는 웨이블렛 변환에 의하여 다중 해상도를 지닌 서브영상들로 분해된다. t 시간의 웨이블렛 신호는 이미 $t-1$ 시간에 처리되어 프레임 메모리에 저장되어 있는 웨이블렛 신호를 기준신호로 사용하여 움직임 추정 및 보상을 수행하여 얻어진 에리신호

DFD(Displacement Frame Difference)와 움직임 벡터를 부호화한다. 이때 DFD는 양자화하여 부호화한다.



<그림 2> 웨이블릿 변환을 이용한 영상 부호화기

3) 블록 매칭 방식에 의한 움직임 추정

움직임 추정에서 가장 많이 쓰이고 있는 블록 매칭 방식은 현재 프레임을 기준 프레임과 블록 단위로 매칭시켜 움직임을 추정하는 기법이다. 즉, 특정 크기의 블록 내 화소(또는 계수)값들이 기준 프레임의 어떤 위치와 가장 잘 일치하는가를 찾아내는 방식이다. 이때 구해지는 움직임 벡터를 수식으로 표현하면 식(2)와 같다.

$$v = \min_{v \in \Omega} \sum_{s \in W} \| I_t(s) - I_{t-1}(s + v) \| \quad (2)$$

여기서, $I_t(s)$ 는 시간 t 에서 $s=(x,y)$ 위치의 블록내 화소값을 의미하며, $v = (v_x, v_y)$ 는 시간 t 일 때의 프레임과 시간 $t-1$ 일 때의 프레임간 움직임 벡터, Ω 는 미리 정해진 탐색 영역, W 는 프레임의 크기에 해당한다.

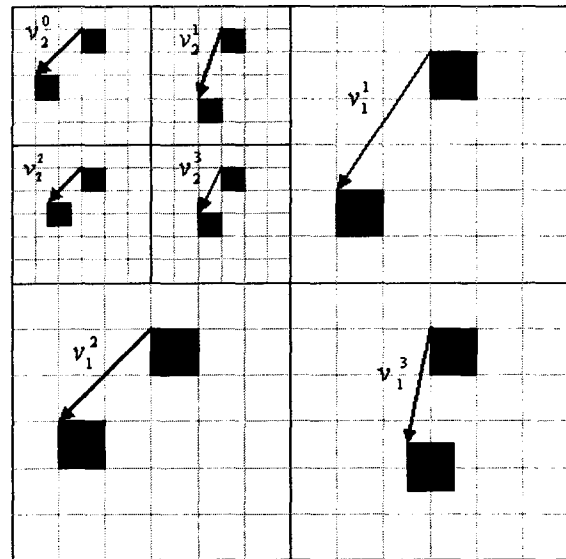
따라서, 식(2)에 의하여 탐색 영역안에서 t 시간의 프레임과 $t-1$ 시간의 프레임간 블록내 화소값의 차인 MAD(Mean of Absolute Difference) 또는 MSE(Mean Square Error)가 가장 작을 때의 위치가 움직임 벡터이다.

움직임 추정시 블록의 크기와 탐색영역은 움직임 추정의 정확도와 추정시간에 매우 중요하게 작용하는 요소가 된다. 만일 탐색영역이 크면 움직임 추정의 정확도는 높아지는 반면, 추정시간이 길게 되며, 탐색영역이 작으면 그 반대의 효과를 가져오게 된다. 따라서, 적합한 탐색영역을 이용하는 것이 중요하다.

II. 기존의 MRME 방식

- 1) 전 화면 탐색 영역 블록 매칭을 이용한 MRME

웨이블릿 피라미드 구조상의 영상들은, 상위 계층과 하위 계층이 상호 연관되어있다. 즉, 상위 계층은 하위 계층에 비하여 저 해상도이며, 영상의 블록크기가 상대적으로 작아야 된다. 만일 m 계층의 웨이블릿 변환을 했다면 j 계층의 해상도는 2^j 이며, 최상위 계층의 블록 크기가 $p \times p$ 이면 j 계층의 블록크기는 $p2^{m-j} \times p2^{m-j}$ 가 되어야 서브영상의 크기를 고려한 다중 해상도 움직임 추정 (MRME; Multiresolution Motion Estimation)이 가능하다.



<그림 3> 전 화면 탐색 영역 블록 매칭을 이용한 MRME

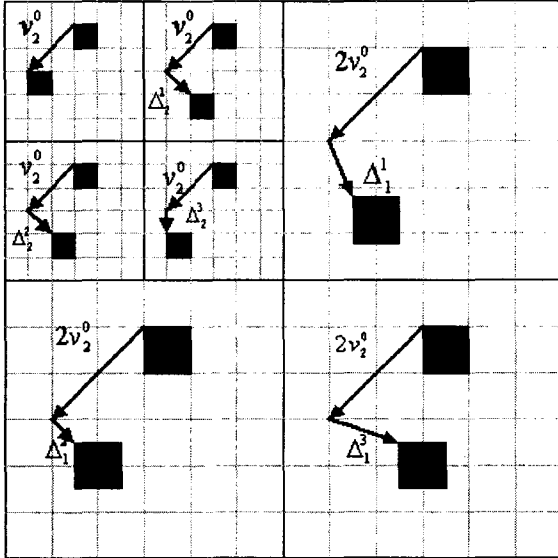
그러나 <그림 3>과 같이 각 계층의 각 서브영상별로 블록 매칭 방식을 사용하여 움직임을 추정하면 에러신호의 크기는 작으나 많은 처리 시간을 요구하며, 웨이블릿이 갖는 계층구조의 장점을 전혀 고려하지 않는 방식이 된다.

2) Zafar 가 제안한 방식

Zhang 과 Zafar 는 웨이블릿 이론을 바탕으로 다양한 블록크기와 해상도의 MRME 방식을 제안하였다[1]. 이 방식은 웨이블릿 피라미드상에서 상위 계층과 하위 계층이 갖는 상관 관계에 의하여 움직임 추정 계산 시간을 감소시킬 수 있다는 장점을 지니고있다. 즉, <그림 4>와 같이 웨이블릿 피라미드의 기초영상에서 움직임 벡터가 추정되면, 이를 기저 벡터로 사용하여 다른 서브영상의 움직임 벡터들을 정제하여 움직임 벡터를 구한다. 웨이블릿이 갖는 계층구조를 이용한 이 방식은 계산 시간을 많이 줄일 수 있다.

이 방식 이외에도, 더 정확한 움직임 추

정을 위해 재귀 알고리즘을 이용하여 상위 계층의 움직임 추정 에러를 수정하려는 방식[1], 더욱 빠른 계산 시간을 위해 적응적인 문턱값을 적용하여 움직임을 추정하고자 하는 AMRME(Adaptive MRME) 방식이 있다[2][3][4].



<그림 4> Zafar 가 제안한 MRME 방식

3) 기존 MRME 방식의 문제점

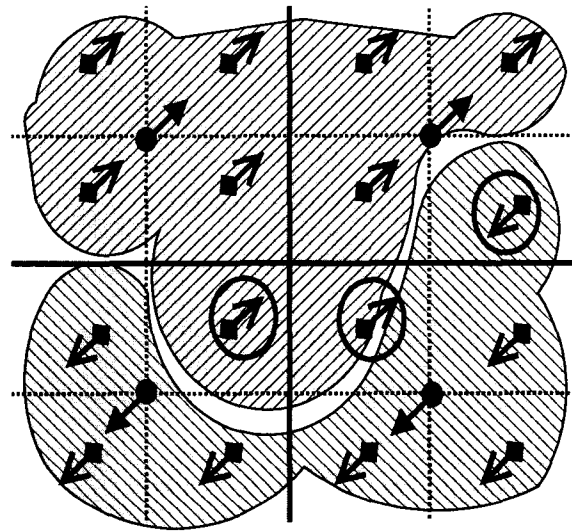
앞에서 설명한 기존의 방식을 이용하여 움직임을 추정하면 기초영상의 블록 크기가 너무 작으므로 틸새(aperture) 문제가 발생하여 실제 영상의 움직임을 나타내는 움직임 벡터를 정확하게 구할 수 없으며 주어진 탐색범위 내에서 단지 에러 신호의 크기가 작은 위치를 나타내는 곳을 움직임 벡터로 찾는다[5]. 따라서 기초영상에서 구한 움직임 벡터를 다른 서브영상의 움직임 벡터를 구하는데 사용하면 다른 서브화면에서 에러신호가 크므로 전체영상에서의 움직임 추정 성능을 저하시킨다.

이러한 문제를 어느 정도 해결할 수 있는 방법으로는 4개의 최하위 계층 영상에서 동시에 움직임을 추정하여 이 결과를 기저 벡터로 사용하는 방법이 있다[2]. 그러나 이 방법은 서브영상간 주파수 성분의 차는 무시하고 영상의 움직임이 같다는 점만을 고려하였으므로 예상보다 성능이 많이 떨어진다.

III. 제안한 방식

기초영상에서 구한 움직임 벡터를 다른 서브영상의 움직임 추정에 사용하려면, 단지 에러신호가 적은 위치만이 아닌 실제로 영상이 움직인 위치를 정확하게 찾아야 한다. 따라서, 틸새문제를 줄이기 위하여 기초영상에서는

부호화할 블록 보다 큰 크기의 블록으로 움직임 벡터를 구한 후 부호화 단계에서는 부호화할 블록크기로 이전 단계에서 구한 움직임 벡터를 기저벡터로 사용하여 정제한다. 계층 수가 많은 경우에는 부호화할 블록의 크기가 너무 작으므로 여러 단계를 거쳐서 움직임을 추정할 필요가 있다. 작은 블록 단계에서의 기저 벡터는 큰 블록 단계에서 구한 자기 위치의 움직임 벡터를 그대로 복제하여 사용하지 않고 자기 위치를 포함하여 이웃한 블록의 움직임 벡터들을 이용하여 작아진 자기 블록내에서 화면간 MAD를 구하여 MAD가 가장 작을 때의 위치를 기저벡터로 사용한다[4].



- : 큰 블록단계에서 움직임 벡터
- ◆→ : 작은 블록단계에서 움직임 벡터
- ⊙→ : 블록 크기에 따라서 움직임 벡터가 다르게 추정되는 경우

<그림 5> 이웃 블록의 움직임 벡터를 이용한 MRME

움직임 벡터는 영상내에서 공간적으로 상관성을 지니므로 DPCM(Differential Pulse Code Modulation) 방식을 이용하여 부호화할 수 있으며 계층적으로 움직임을 추정하면 주위 블록과의 움직임 벡터 편차가 작아지므로 비트량을 감소시킬 수 있다.

기초영상을 제외한 다른 서브영상에서는 움직임을 추정하기 전에 기초영상에서 구한 움직임 벡터를 이용하여 MAD를 구한 후 MAD가 미리 정해진 문턱값 보다 작을 경우에는 움직임 추정 과정을 생략하고 기초영상에서 구한 움직임 벡터를 그대로 이용하면 계산시간을 절약할 수 있다. 또한 여러개의 문턱값을 두고서 MAD의 크기에 따라서 가변적인

탐색영역을 사용할 수 있다.

IV. 실험 및 결과

사용한 실험 데이터는 움직임 특성이 서로 다른 Football 과 Flower garden 을 이용하여 실험하였다.

먼저, 각 영상을 2 계층 2D-DWT 하여 생성된 서브영상의 계수들을 이용하여 기존의 방식들과 본 논문에서 제안한 방식으로 MRME 를 수행하였다. 각 방식에 대하여 PSNR(Peak signal-noise ratio), DFD 의 엔트로피, 움직임 벡터의 엔트로피 그리고 움직임 추정 계산시간을 구하여 <표 1>에 나타내었다.

실험한 방식들은 다음과 같다.

- i. 방식 I : 계층별로 화면의 크기에 비례한 탐색영역을 사용한 전 화면 탐색 블록 매칭
- ii. 방식 II : 기초영상의 움직임 벡터를 다른 서브영상에서 기저벡터로 사용하는 방식
- iii. 방식 III : 4 개의 최상위 계층 서브영상을 동시에 움직임 추정하여 얻어진 움직임 벡터를 각 서브영상에서 기저벡터로 사용하는 방식
- iv. 방식 IV-1 : 본 논문에서 제안한 방식(문턱값을 사용하지 않은 경우)
- v. 방식 IV-2 : 본 논문에서 제안한 방식(문턱값을 사용한 경우)

<표 1> 실험 결과

	PSNR (dB)	Error image Entropy (bit/pixel)	Motion Vector Entropy (bit/vector)	* Processing Time (sec)
"Flower garden" sequence				
방식 I	25.32	4.36	6.28	44.49
방식 II	25.15	4.37	6.52	15.11
방식 III	24.89	4.38	6.71	19.66
방식 IV-1	25.36	4.36	6.40	16.37
** 방식 IV-2	25.35	4.37	6.11	13.57
"Football sequence" sequence				
방식 I	26.11	4.12	5.70	44.71
방식 II	25.89	4.13	5.61	14.99
방식 III	25.86	4.15	6.01	19.61
방식 IV-1	25.93	4.13	5.60	16.37
** 방식 IV-2	25.92	4.13	5.51	14.72

* 펜티엄-II CPU 266MHz PC 에서 1 프레임의 영상을 처리하는데 걸린 평균시간

** 문턱값: 3

본 논문에서 제안한 방식의 PSNR 이 방식 II 나 방식 III 에 비하여 좋음을 알 수 있으며, 방식 I 과 비교하면 Flower garden 에서는 제안한 방식이 좋으며 Football 에서는 방식 I 이 좋음을 알 수 있다.

제안한 방식에서 문턱값을 사용하면 엔트로피가 향상되었으며 실험한 방식 중에서 가장 우수하였다. 또한 처리시간은 16%정도 절약한 반면에 PSNR 은 0.01%이하 밖에 감소되지 않았다.

<그림 6>, <그림 7> 그리고<그림 8>에 실험에 사용된 Flower garden 영상과 움직임 보상된 영상 그리고 에러신호를 나타내었다. <그림 8>의 아랫부분에서 알 수 있듯이 제안한 방식에 의하여 실험한 결과가 기존의 방식에 비하여 에러성분이 훨씬 적음을 알 수 있다.

VI. 결 론

웨이블렛으로 부호화된 신호들은 여러 종류의 영상 포맷 사이에서 유동적으로 사용되어질 수 있으며 블록킹 효과가 적으므로 앞으로 영상처리에 많이 이용될 예정이다. 본 연구는 웨이블렛 부호화기에서 화면간 영상 데이터의 압축을 위하여 제안하였으며 기존의 방식과 실험을 통하여 비교 분석하였다.

좋은 MRME 방식이란 정확한 움직임 추정과 적은 계산시간이 가능하여야 한다. 본 논문에서 제안한 MRME 방식은 기초영상에서 계층적 블록 매칭을 이용하여 실제적인 움직임 벡터를 구한 후 이를 다른 서브영상에서도 이를 이용하여 전체영상에서의 움직임 추정 성능을 향상시켰으며, 기존의 MRME 방식 보

다 엔트로피와 처리시간 면에서 우수함을 확인하였으며 특히 문턱값을 적용한 경우에는

엔트로피와 처리시간 면에서 가장 우수함을 확인하였다.



a) 현재영상 b) 기준영상 c) 두 영상간의 차 (x3)

<그림 6> 2-D DWT 한 후 2 장의 기초영상과 영상간의 차



a) 방식 I 및 방식 II b) 방식 III c) 방식 IV

<그림 7> MRME 한 결과를 이용하여 움직임 보상한 영상



a) 방식 I 및 방식 II b) 방식 III c) 방식 IV

<그림 8> 움직임 보상한 영상과 원영상과의 차 (x3)

참고문헌

[1] Sohail Zafar, "Multiscale Video Representation using Multiresolution Motion Compensation Wavelet Decomposition", *IEEE journal on selected areas in communications*, VOL. 11, NO. 1, pp. 24-35, JANUARY 1993.

[2] Seongman Kim, Seunghyeon Rhee, "Interframe Coding using Two-Stage Variable Block-size Multiresolution Motion Estimation and Wavelet Decomposition", *IEEE Transactions on circuits and systems for video technology*, VOL. 8, NO. 4, pp. 399-410, AUGUST 1998.

[3] Ya-Qin Zhang, "Motion-Compensated Wavelet Transform Coding for Color Video

Compression", *IEEE Transactions on circuits and systems for video technology*, VOL. 2, NO. 3, pp. 285-296, SEPTEMBER 1992.

[4] Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: A review and a new contribution", *Proc. IEEE*, VOL. 83, pp. 856-876, JUNE 1995.

[5] Murat Tekalp, "Digital Video Processing", Prentice Hall, pp.78-79, 1995

<감사의 글>

본 논문이 완성되기까지 도움을 주신 ETRI의 방송기술연구부원과 ICU의 멀티미디어 그룹 학생들에게 감사드립니다.