

# 다 해상도 프레임 구조에 기반한 고속 움직임 추정 기법

송병철, 나종범

한국과학기술원 전기 및 전자공학부

Tel: 042-869-5434, Fax: 042-869-8360

E-mail: bcsong@issserver.kaist.ac.kr, jbra@ee.kaist.ac.kr

## A FAST MOTION ESTIMATION ALGORITHM BASED ON MULTI-RESOLUTION FRAME STRUCTURE

Byung Cheol Song and Jong Beom Ra

Department of Electrical Engineering, KAIST

Tel: 042-869-5434, Fax: 042-869-8360

E-mail: bcsong@issserver.kaist.ac.kr, jbra@ee.kaist.ac.kr

### ABSTRACT

We present a novel multi-resolution block matching algorithm (BMA) for fast motion estimation. At the coarsest level, a full search BMA (FSBMA) is performed for searching complex or random motion. Concurrently, spatial correlation of motion vector (MV) field is used for searching continuous motion. Here we present an efficient method for searching full resolution MVs without MV decimation even at the coarsest level. After the coarsest level search, two or three initial MV candidates are chosen for the next level. At the further levels, the MV candidates are refined within much smaller search areas. Simulation results show that in comparison with FSBMA, the proposed BMA achieves a speed-up factor over 710 with minor PSNR degradation of 0.2dB at most, under a normal MPEG2 coding environment. Furthermore, our scheme is also suitable for hardware implementation due to regular data-flow.

### 1. 서론

MPEG2 나 H.261 같은 동영상의 부호화에서 연속된 프레임들간의 시간적 중복성을 제거하기 위해 움직임 보상 예측 부호화 기법이 널리 사용되고 있다 [1-2]. 블록 정합 기법 (BMA) 은 대표적인 움직임 보상 기법으로서, 이전 프레임의 탐색 영역 내에서 현재 프레임의 기준 블록과 가장 유사한 블록을 찾는 방식이다. 특히 전역 탐색 기법 (FSBMA) 은 탐색 영역 내의 모든 블록들을 탐색하기 때문에 최적의 방법이라고 할 수 있다. 그러나 FSBMA 는 계산량이 방대하여 넓은 탐색 영역을 요구하는 MPEG2 나 HDTV 같은 영상 시스템에는 적합하지 않다. FSBMA 의 방대한 계산량을 줄이기 위해 수많은 고속 BMA 들이 개발되었다 [3-11].

고속 BMA 들은 특성에 따라 다음과 같이 몇 가지로 분류할 수 있다. 단조로운 오류 표면 모델에 기반한 LOGS 나 TSS 같은 BMA [3-4], 화소 추출 (pixel subsampling) 을 통한 BMA [5], 다 해상도 BMA [6-8], 그리

고 시간적/공간적 상관성을 이용한 BMA [9-10]. 첫 번째 BMA 는 이 기본 모델이 실제 영상에 잘 맞지 않기 때문에 국소적 최소치 (local minimum) 에 빠지기 쉽다. 두 번째 BMA 는 계산량 감소를 위해 과도하게 화소 추출을 할 경우 역시 국소적 최소치 문제기 생기기므로 계산량 감소에 한계가 있다. 세 번째 종류의 BMA 는 비교적 좋은 성능과 빠른 계산 속도를 갖는다. 그러나 대개의 경우 시간적/공간적인 상관성 같은 부가 정보를 이용하지 않고, 이전 계층의 움직임 정보만을 이용하여 현재 계층의 움직임을 추정하기 때문에 상당 밀을 만한 성능을 보이는 것은 아니다. 마지막 종류의 BMA 는 예측 탐색에 기반한 기법이다. 이런 기법은 시간적/공간적 상관성을 이용함으로써 계산량을 줄일 뿐만 아니라 상관성이 강한 움직임 벡터 장 (MV field) 을 갖는 영상들에 대해서는 상당히 밀을 만한 성능을 보인다. 그러나 이런 기법들은 기본적으로 움직임 벡터 장이 강한 상관성을 갖는다고 가정하기 때문에 가정이 맞지 않는 영상에서는 성능 저하가 심각해진다. 또한 불규칙적인 데이터 흐름 (data-flow) 으로 하드웨어 구현에 적합하지 않다.

위에서 언급한 기존 고속 BMA 들의 단점을 극복하기 위해, 본 논문에서는 공간적으로 인접한 움직임 벡터들간의 상관성을 이용한 새로운 다 해상도 BMA 를 제안한다. 각 계층에서 SAD (sum-of-absolute-difference) 를 계산할 때 규칙적인 데이터 흐름을 유지하기 위해 효율적인 SAD 예측 기법을 제안한다. 상황에 따라 기준 블록의 SAD 대신 기준 블록과 공간적으로 약간 이긋난 블록의 SAD 를 사용하는 것이다. 이런 예측 기법은 다음과 같은 두 가지 장점을 지닌다. 하나는 이전 프레임 데이터의 규칙적인 처리로 인한 하드웨어 구현의 적합성이며, 다른 하나는 낮은 해상도 계층에서도 움직임 벡터의 양자화 없이 움직임 추정이 가능하다는 점이다. 제안한 SAD 예측 기법의 정확성이 3개의 모의 실험을 통해 증명된다. 이 방법으로 각 계층에서의 움직임 벡터 추정을 정수 단위로 수행할 수 있으며, 최상 해상도 계층에서의 지역 탐색을 성능 저하 없이 생략하여 추가

적으로 계산량을 줄일 수 있다. 따라서 제안한 BMA 는 빠른 속도뿐만 아니라 좋은 성능을 갖는다. 또한 구조의 단순성과 규칙적인 데이터 흐름으로 하드웨어 구현에 적합하다.

본 논문은 다음과 같이 구성된다: 2절에서 새로운 다 해상도 BMA 를 제안하며, 3 절에서는 모의실험 결과를 보인다. 마지막으로 4 절에서 결론을 맺는다.

## 2. 고속 다 해상도 블록 정합 기법

그 동안 움직임 추정을 비롯한 여러 응용 분야에 다 해상도 기법을 적용하는 연구가 활발히 진행되어 왔다. 다 해상도 BMA 들에서는 각 프레임이 여러 해상도를 갖는 몇 개의 계층으로 나뉜다. 계층 별로 움직임 벡터 정보를 가지며, 이전 계층에서 얻어진 몇 개의 움직임 벡터 후보들을 중심으로 현재 계층에서 보다 세밀한 움직임 추정을 한다. 최근에 다 해상도 구조와 시간적/공간적 상관성을 이용한 고속 BMA 가 개발된 바 있다 [8]. 이 기법은 다른 고속 BMA 들에 비해 좋은 성능을 보이지만 다음과 같은 두 가지의 단점을 갖는다. 첫째, 불규칙한 데이터 흐름으로 하드웨어 구현에 적합하지 않다. 둘째, 복잡한 움직임을 갖는 영상들에 대해서는 좋지 않은 성능을 보인다. 이것은 낮은 해상도 계층에서의 움직임 추정 시 후보 벡터로 쓰이는 이전 프레임의 움직임 벡터들이 계층에 맞게 양자화 된 후 사용되어 정확성이 상당히 떨어지기 때문이다.

본 절에서는 MRBMA 라는 새로운 다 해상도 BMA 를 제안한다. MRBMA 는 빠른 속도 및 최상의 성능을 보인다. 또한 하드웨어 구현에도 적합하다.

### 2.1 MRBMA 의 프레임 구조

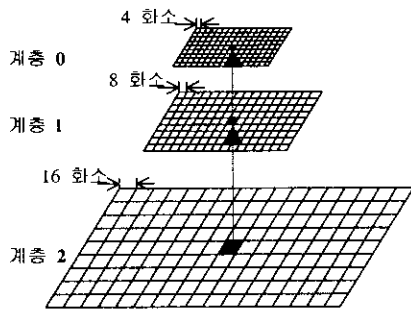


그림 1.  $L=3$  인 경우의 다 해상도 프레임 구조. 각 사각형은 한 매크로블록 (MB) 을 의미한다.

본 논문에서는 편의상 프레임 기반 움직임 추정만을 고려한다.  $L$  은 계층 수를 의미하며, 0 가 최하 해상도 계층을,  $L-1$  이 최상 해상도 계층을 가리킨다.

MRBMA 에서 이전 프레임을 위한 데이터 구조로 다 해상도 구조가 이용된다. 그림 1 에서 알 수 있듯이, 낮은 해상도 계층의 각 화소는 이전 계층의  $2 \times 2$  블록의 평균을 구함으로써 얻어진다.

한편 MRBMA 는 그림 2 처럼 현재 프레임의 구조로서

블록 평균 피라미드 구조를 선택한다 [11]. 다 해상도 프레임 구조와 달리, 블록 평균 피라미드의 프레임 및 MB 크기가 각 계층에서 일정하다.

현재 프레임 구조 형성을 위한 계산량을 살펴보면 화소 당 2 회의 정합 연산에 해당하는 3 회의 덧셈과 1 회의 이동 (shift) 연산이 요구된다. 따라서 전체 움직임 추정을 위한 계산량과 비교할 때 무시할 만하다. 한편, 다 해상도 프레임 구조를 위한 계산량은 이에 비해 훨씬 적다.

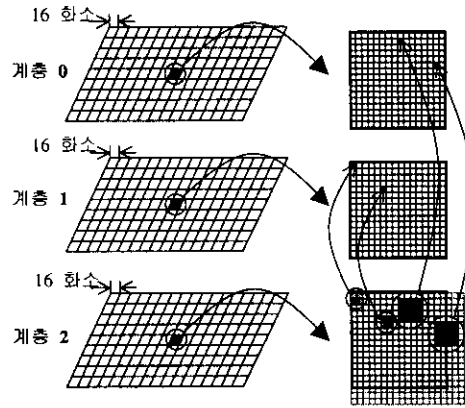


그림 2. 현재 프레임을 위한 블록 평균 피라미드 프레임 구조 ( $L=3$ ). 각 사각형은 한 MB 을 의미한다.

### 2.2 MRBMA 의 프레임워크

MRBMA 는 [8]의 BMA 처럼 계층 0 에서 FSBMA 를 수행한다. 계층  $L-1$  의 관점에서 계층 0 의 FSBMA 는 전체 탐색 영역  $\Omega_s ((2N_s+1) \times (2N_s+1))$  내에서 규칙적으로 추출된 움직임 벡터들을 탐색하는 것과 동일하다. 보통 이런 탐색은 복잡한 움직임을 추정하는데 유리하다. 계층 0 의 FSBMA 에서 기준 MB  $B_{(i,j)}$  (우측 최상단의 좌표가  $(i,j)$  ) 의 주변 블록들의 움직임 벡터들 ( $V_m^n, 1 \leq m \leq 4$  ) 이 초기 움직임 벡터 후보로서 탐색된다 (그림 3 참조). 이  $V_m^n$  들은 주로 연속적인 움직임을 추정하는데 유리하다 [10]. MRBMA 의 첫번째 중요한 특징이 바로 움직임 벡터의 양자화 없이 계층 0 에서도 네 개의  $V_m^n$  들을 탐색할 수 있다는 점이다.

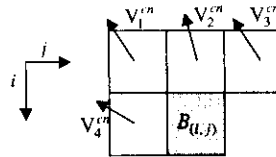


그림 3. 주변 MB 들의 기하학적 위치.

이는 MRBMA 가 위에서 언급한 프레임 구조들을 가지기 때문에 가능하다. 기존의 다 해상도 BMA 에서의 움직임

직임 벡터의 양자화 없이 계층 0에서  $V_m^{cn}$  들을 탐색하는 것이 불가능했거나,  $V_m^{cn}$  들의 계층 0에서의 탐색 자체가 무의미했다 [8].

MRBMA에서는  $V_m^{cn}$  에 대응하는  $B_{(i,j)}$  의 SAD를 예측하기 위해 약간 이동된 블록  $B_{(i+\alpha, j+\beta)}$  의 SAD를 이용한다 (그림 4 참조). 그림 4에서 이전 프레임의 화소 좌표들이 편의 상 현재 프레임의 관점에서 기술된다. 만약,  $V_m^{cn}$  의 좌표가  $(p, q)$  라면 계층 1에서의  $\alpha$  와  $\beta$  는 다음과 같이 정의된다.

$$\alpha = \left\lfloor \frac{p}{2^{L-1}} \right\rfloor \cdot 2^{L-1} - p, \quad \beta = \left\lfloor \frac{q}{2^{L-1}} \right\rfloor \cdot 2^{L-1} - q \quad (1)$$

(1)에서,  $\lceil A \rceil$  는 A 보다 크거나 같은 가장 작은 정수들의 의미이다.

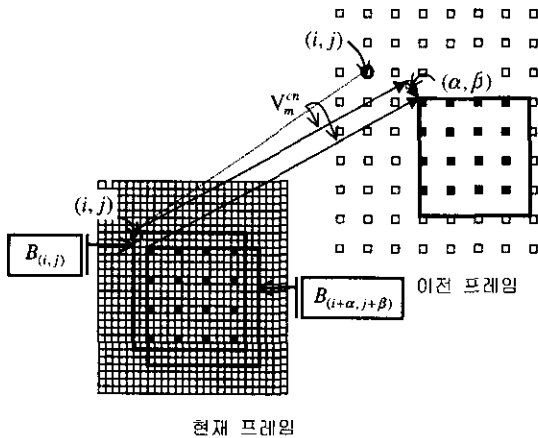


그림 4. L=3인 경우의 계층 0에서의  $V_m^{cn}$  탐색. SAD 계산 시 검은 색 화소들만이 이용된다.

제안한 SAD 예측 기법은 움직임 벡터의 정확성이 보장되고  $\alpha$  와  $\beta$  의 값이 그리 크지 않아  $B_{(i,j)}$  내의 대부분의 화소들이  $B_{(i+\alpha, j+\beta)}$  에도 포함되기 때문에 상당한 정확성을 갖는 예측이 가능하다.

FSBMA 동안 양자화된  $V_m^{cn}$  의 탐색은 생략된다. 대신에, 그 위치에서 제안한 SAD 예측이 수행된다. 이는  $V_m^{cn}$  이 양자화된  $V_m^{cn}$  에 비해 영역 내 최소 (global minimum) 일 가능성이 클 뿐만 아니라 이전 프레임의 데이터 흐름의 규칙성을 유지 할 수 있기 때문이다.

계층 0에서의 FSBMA를 수행한 후에 계층 1에서 여러 개의 지역 탐색을 수행한다. 그러나 지역 탐색의 수가 많아지면 하드웨어 구현이 어려워지고 복잡도가 증가하기 때문에 계층 0에서 SAD가 작은 두 개의 후보만을 선택한다. 첫 번째 후보  $V_1^{L0}$  는 네 개의  $V_m^{cn}$  들 중에서 선택되며, 다른 하나  $V_2^{L0}$  는 나머지 움직임 벡터 후보 중에서 선택한다. 여기서 위치자 0는 계층 0

를 의미한다. 마지막으로, 두 초기 벡터를 중심으로 하는 지역 탐색을 수행하여 계층 2를 위한 초기 벡터  $V^{L1}$  을 찾는다. 탐색 영역  $\Omega_{L1}$  는  $(2N_{L1}+1) \times (2N_{L1}+1)$  으로 고정한다 (본 실험에서는  $3 \times 3$  이나  $5 \times 5$ ). 이 탐색 영역은  $\Omega_5$  에 비해 현저하게 줄어든 것이다. 기존의 다해상도 BMA 들에서는 계층 L-1 에서만 정수 화소 단위의 지역 탐색이 가능했던 반면, MRBMA 의 각 계층 탐색은 제안한 SAD 예측 기법으로 정수 화소 단위로 수행되어짐에 주목해야 한다. 바로 이 점이 MRBMA 의 또 다른 특징이다. 따라서 계층 L-1 에서의 지역 탐색은 MRBMA 의 전체 성능에 미미한 영향을 주어 성능 저하 없이 계층 L-1 에서의 탐색을 생략하여 추가적인 계산량 감소를 얻을 수 있다. 3절에서의 모의 실험 결과가 이를 뒷받침해준다.

계층 2에서,  $V^{L1}$  중심의 지역 탐색 후 최종적인 움직임 벡터  $V^{L2}$  을 얻는다. 여기서 탐색 영역  $\Omega_{L2}$  는  $(2N_{L2}+1) \times (2N_{L2}+1)$  이다 (우리 실험에서는  $3 \times 3$ ). 위에서 언급 했듯이, 이 마지막 탐색은 생략이 가능하다.

### 3. 모의실험 결과

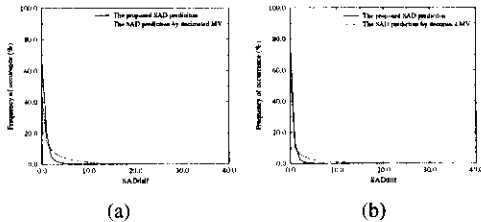
실험을 위해, 다음 MPEG2 동영상들의 처음 100 프레임이 사용되었다; "cheer leaders (cheer)," "car (car)," "football (foot)," "flower garden (flower)." 각 프레임은  $720 \times 480$  의 해상도를 갖는다. MRBMA-mLnS 는 m 계층을 갖는 n-단계 MRBMA 를 의미한다. 보통 n 은 m 이나 m-1 이 된다. 예를 들면, MRBMA-3L2S 는 L=3 인 경우 마지막 계층 탐색이 생략된 MRBMA 를 의미한다.

#### 3.1 제안한 SAD 예측 방법의 정확도

먼저 계층 2에서 FSBMA 를 통해 각 MB 의 정확한 움직임 벡터를 찾는다. 그런 다음 계층 0에서 두 종류의 SAD 를 계산한다. 제안한 예측 기법에 의한 SAD 과 움직임 벡터 양자화에 의한 SAD 가 그것이다. 마지막으로 두 SAD 들을 계층 0에서 인위적으로 구한 정확한 SAD 와 비교한다. 그림 5 의 (a) 는 그 실험 결과를 보여준다. 그림 5 에서,  $SAD_{diff}$  는 두 SAD 의 절대차이며, 256 (MB 크기) 로 나누어진 값이다. 좀 더 명확히 보기 위해, 그래프들은  $SAD_{diff}$  이 0 부터 40 까지 도시 된다. 그림 5 의 (b) 는 계층 1에서의 결과를 보여준다. 위 결과들로부터 제안한 SAD 예측 기법이 매우 신뢰성이 있으며, 움직임 추정에 사용하기에 충분히 좋은 성능을 갖는다는 것을 확인할 수 있었다.

#### 3.2 MPEG-2 동영상 부호화기에서의 MRBMA 의 성능 평가

GOP 크기는 15 프레임, P 프레임 간 거리는 3 프레임, 비트 율은 4 Mbps 로 고정된 후 실험하였다.  $\Omega_5$  는 P 프레임의 경우  $127 \times 127$  이며, B 프레임의 경우  $63 \times 63$  로 정했다. MRBMA 의 성능은 평균 PSNR ( $PSNR_{avg}$ ) 전지에서 평가되며, 표 1 에 그 결과가 나타나 있다.



(a) (b)  
 그림 5. L=3인 경우 제한한 SAD 예측기법의 정확도 테스트 (a) 계층 0 (b) 계층 1.

MRBMA가 FSBMA와 거의 같은 PSNR<sub>avg</sub>을 가짐을 알 수 있다. 특히 지역 탐색 영역의 크기에 관계없이 “flower garden”과 “cheerleaders” 두 영상에 대해 FSBMA와 동등한 성능을 보이는 점은 주목할 만하다.

다양한 탐색 영역에 대해 MRBMA의 복잡도를 표 II에 기술하였다. MB 당 연산 수를 기준으로 하여 FSBMA과의 상대적인 복잡도를 속도향상도라고 정의하고 복잡도의 척도로 사용하였다.

위 표에서 MRBMA가 상당한 복잡도의 감소와 함께 FSBMA와 거의 같은 성능을 보임을 확인할 수 있었다.

표 I. MRBMA와 FSBMA간의 평균 PSNR 비교.

기법	동영상 시퀀스				
	car	cheer	foot	flower	
FSBMA	35.3	28.9	31.5	29.5	
MRBMA	3L3S ( $N_{L1}=2, N_{L2}=1$ )	35.2	28.9	31.5	29.5
	3L3S ( $N_{L1}=N_{L2}=1$ )	35.1	28.9	31.5	29.5
	3L2S ( $N_{L1}=2$ )	35.1	28.9	31.5	29.5
	3L2S ( $N_{L1}=1$ )	35.1	28.9	31.4	29.5
	4L3S ( $N_{L1}=N_{L2}=1$ )	35.1	28.9	31.4	29.5

표 II. 속도향상도에 따른 복잡도 비교.

MRBMA	속도향상도				
	$N_8=7$	$N_5=15$	$N_3=31$	$N_2=63$	$N_1=127$
3L3S ( $N_{L1}=2, N_{L2}=1$ )	7	27	83	169	227
3L3S ( $N_{L1}=N_{L2}=1$ )	11	40	111	194	237
3L2S ( $N_{L1}=2$ )	11	43	116	197	238
3L2S ( $N_{L1}=1$ )	29	92	178	231	249
4L3S ( $N_{L1}=N_{L2}=1$ )	24	101	387	1222	2585

#### 4. 결론

본 논문에서는 넓은 탐색 영역에 대해서도 매우 빠른 속도와 좋은 성능을 갖는 새로운 다 해상도 BMA (MRBMA)를 제안한다. 모의실험을 통해 MRBMA가 127 x 127의 넓은 탐색 영역에 대해서도 좋은 성능을 보임을 확인했다. 예를 들면, MRBMA-4L3S는 FSBMA의 0.14%에 해당하는 계산량만으로 최악의 경우 0.2dB

라는 미미한 PSNR 저하만을 보인다. 따라서, MRBMA는 실시간 동영상 부호화에 유용하게 이용될 수 있다. 또한 MRBMA는 전체 탐색 과정동안 규칙적인 데이터 흐름을 갖기 때문에 하드웨어 구현에도 적합하다.

#### 5. 참고 문헌

- [1] MPEG2, “Information technology-generic coding of moving pictures and associated audio,” Tech. Rep., ISO/IEC 13818-2, Committee Draft, Mar. 1994.
- [2] CCITT Study group XV, “Draft revision of recommendation H.261-Video codec for audio visual services at px64 kbps,” Temporary Document 5-E, July 1990.
- [3] J. R. Jain and A. K. Jain, “Displacement measurement and its application in inter-frame image coding,” *IEEE Trans. Commun.*, vol. COM-29, pp. 730-741, 1981.
- [4] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, “Motion compensated inter-frame coding for video conferencing,” *Proc. Nat. Telecommunication Conf.*, pp. G5.3.1-5.3.5, Nov. 29-Dec. 3, 1981.
- [5] B. Liu and A. Zaccarin, “New fast algorithms for the estimation of block motion vectors,” *IEEE Trans. Circ. Sys. Video Technol.*, vol. 3, pp. 148-157, Apr. 1993.
- [6] M. Bierling, “Displacement estimation by hierarchical block matching,” *Proc. SPIE Visual Communications and Image Processing '88*, vol. 1001, pp. 942-951.
- [7] K. M. Nam, J. S. Kim, R. H. Park, and Y. S. Shim, “A fast hierarchical motion vector estimation algorithm using mean pyramid,” *IEEE Trans. Circ. Sys. Video Technol.*, vol. 5, no. 4, pp. 344-351, Aug. 1995.
- [8] J. Chalidabhongse and C.-C. J. Kuo, “Fast motion vector estimation using multiresolution-spatio-temporal correlations,” *IEEE Trans. Circ. Sys. Video Technol.*, vol. 7, no. 3, pp. 477-488, June 1997.
- [9] S. Zafar, Y. Q. Zhang, and J. S. Baras, “Predictive block matching motion estimation for TV coding—part I: Inter-block prediction,” *IEEE Trans. Broadcasting*, vol. 37, no. 3, Sep. 1991.
- [10] B. C. Song and J. B. Ra, “Hierarchical block matching algorithm using partial distortion criterion,” *Proc. SPIE Visual Communications and Image Processing*, vol. 3309, 1998.
- [11] W. Li and E. Salari, “Successive elimination algorithm for motion estimation,” *IEEE Trans. Image Processing*, vol. 4, no. 1, Jan. 1995.