

영역 분할 정보를 이용한 고속 움직임

추정 부호화 및 의미 영역 인식

^{*)}이 봉호, 서 정구, 광 노윤, 강 태하, 황 병원

한국항공대학교 전자공학과

A

Fast Motion Estimation Coding and Semantic Region Recognition Using Segmented Region Information

^{*)}Bong-Ho Lee, Jung-Gu Seo, No-Yoon Kwak, Tea-Ha Kang, Byung-Won Hwang

Dept. of Avionics, Hankuk Hangkong Univ.

Email : bhlee@mail.hangkong.ac.kr

요 약

본 논문에서는 초저속 동영상 부호화에 관한 것으로, 움직임 추정 효율을 개선하기 위해 분할된 영역별로 움직임 정보를 추정하여 부호화를 수행할 뿐만 아니라 분할된 영역중 의미있는 부분을 선택적으로 부호화할 수 있는 영역 분할 기반 영상 부호화 기법에 관한 것이다.

첫째로, 움직임 추정은 분할된 영역 정보를 이용한 가변 탐색 영역 설정을 통해 전역 탐색 움직임 추정 시 소모되는 많은 연산량을 줄이고, 둘째로, 움직임 추정 후 추정된 움직임 정보를 이용해 영역의 재분할 과정을 통해 분할된 영역별로 움직임 정보를 부호화 함으로써 개선된 부호화 효율을 보이며, 셋째로, 분할된 영역 중 얼굴과 같은 의미를 갖는 영역을 선택적으로 부호화하고 전송하기 위한 인식 기법을 제안하고자 한다.

I. 서 론

디지털 이동 통신(GSM)이나 PSTN을 이용한 비디오 데이터의 전송은 매우 제한적인 대역을 갖음으로 인해서 64kbps이하의 데이터 압축이 요구되는 실정이며 이를 위한 많은 영상 압축 기법이 연구 중이다. 현행 ITU-T의 H.26p 계열의 압축 표준안들은 DCT를 기반으로 이루어지는 상태이고 보다 나은 부호화 효율을 위해서 영상에 의미를 부여하여 부호화하는 의미론 기반 부호화(Semantic-based coding), 영역 기반 부호화(Region-based coding), 특정한 모델 설정을 통한 모델 기반 부호화(Model-based coding), 객체기반 부호화(Object-based coding) 등이 연구되어 지고 있는 상태이다. MPEG-4에서는 객체(Object) 추출을 통한 VOP(Video Object Plane)를 형성해 VOP별로 부호화한다. 본 논문에서는 H.26p에 분할기반 부호화(Segment-based coding)기법을 이용해 성능 개선을 보이고자한다.

몇 가지 개선해야될 문제점을 살펴보면, 우선 움직임 추정은 부호화 성능에 있어서 매우 중요한 기법으로서 움직임 추정이 잘 이루어지면 이루어질수록 여러 이미지의 부호화 비트율이 낮아지므로 H.26p에서는 움직임

추정을 위한 탐색 방법으로 전역 탐색 기법(FSBMA: Full Search Block Matching Algorithm) 및 TSS(Three Step Search) 기법 사용을 주로 권고하고 있지만 이들 기법은 탐색 시 계산량이 많은 관계로 실시간 응용 시 많은 어려움이 있다.[1] 이를 위해 본 논문에서는 영역 분할 정보를 이용해 구성된 영역별로 탐색 영역을 가변적으로 적용해 계산량을 줄이는 분할 기반 움직임 추정(Segment-based Motion Estimation)기법을 제안하고자 한다. 다음으로는 현행 H.263에서는 B-Picture를 부호화하는데 있어서 움직임 정보만을 보냄으로써 큰 부호화 효율을 얻고 있는데 영역 분할 정보를 이용해 B-Picture 전송 시 부호화 비트율을 낮추는 기법을 제안하며, 마지막으로 초 저속용 비디오 부호화는 많은 가변 분해능(Scalability)이 요구되는데 H.26p 기반으로 분할된 영역의 분석을 통한 얼굴이나 손 같은 의미 있는 영역의 선택적인 부호화를 지원하며 이를 선택적으로 전송하기 위한 인식 기법을 제안하고자 한다. 얼굴영역은 색상값 및 얼굴 영역이 일반적으로 위치하는 부분을 윈도우(Window)로 설정해서 분석한다.

본 논문의 구성은 다음과 같다. II장에서는 분할 기반 움직임 추정(Segmented-based Motion Estimation)기법을 기술하며, III장에서는 B-Picture의 영역별 움직임 정보 부호화를 제안하고, IV장에서는 의미를 갖는 영역의 우선적 전송을 위한 인식 기법을 제안하며, V장에서는 실험 결과 및 검토 과정을 거치며, 마지막으로, VI에서는 결론과정을 끝으로 본 논문을 마무리한다.

II. 분할기반 움직임 추정

기존의 전역탐색 움직임 추정 기법은 각 매크로 블록별로 ±16화소 정도의 탐색 영역(Search Window)을 설정해 움직임을 추정하였으나 화상 회의(Video Conferencing)나 화상전화(Videophone)의 경우에 있어서 통계적으로 적은 움직임을 갖는 특성이 있어 ±16범위 움직임 추정을 하는 데에는 비효율적인 면이 있고 계산상에도 많은 계산 시간이 필요하다.[2-3] 고속 압도 리즘으로 3단계 탐색 기법(TSS) 및 계층적 움직임 추정

기법(HBMA)등이 연구되었지만 본 논문에서는 움직임을 갖는 영역들을 대략적으로 분할한 후, 이 분할 정보를 이용해 분할된 영역 중 임의로 몇 블록을 추출해 ± 16 까지 움직임을 추정해 그 영역이 갖는 움직임 범위를 가변적으로 설정해 고속 탐색을 실현하였다. 움직임 추정과정은 다음과 같다.

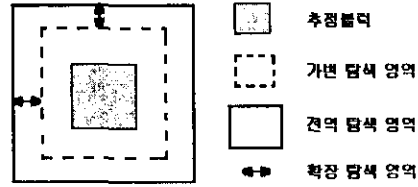


그림 1. 확장 추정 영역

다. 영역 재병합(Region Re-Merging)은 움직임 추정을 위해서 전 단계에서 구성된 분할 영역에 추정된 움직임 벡터를 이용해 재구성하게 된다. 재병합 과정은 다음과 같다.

1. 가우시안 필터링 : 두 영상에 대해서 움직임의 상관을 개선하기 위해 3x3 가우시안 필터링을 취한다.
2. 두 프레임의 차 영상(Difference Image)을 구한다.
3. 영역 병합 기법을 이용한 영역 분할 : 차 영상에 대해 8x8 블록으로 나눈후, 임계치(Th1)이하의 차 화소수를 가진 블록은 제거하고 영역 병합(Region Merging)기법을 이용해 움직임을 갖는 영역을 구성한다. 병합 시 8 방향을 고려한다.[4] 이 과정에서 마커(Marker)를 추출한다.

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

RegionMerging();

MarkerExtraction();

여기서, $M(B_i)$ 는 블록 i의 병합 조건

병합시 차 영상에 나타나는 고립된 작은 블록들은 잡음에 의한 블록에 가깝고 그다지 영상에서 의미를 갖지 않으므로 제거한다.

4. 샘플 블록의 추출 및 움직임 추정 : 분할된 영역들의 블록 수에 비례하여 샘플 블록을 임의로 추출한 후, ± 16 범위로 움직임 추정 상관함수인 MAD(Mean Absolute Difference)를 이용해 움직임을 추정한다.

$$MAD(i, j) = \frac{1}{M \times N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |I_i(x+m, y+n) - I_{j-1}(x+m+i, y+n)|$$

5. 가변 탐색 영역 결정 : 영역별로 추정된 움직임 벡터들 중 가장 값이 큰 벡터들 부터 신뢰도를 측정한 후 가변 탐색 영역을 설정한다. 신뢰도 측정 과정은 구성된 블록들의 평균 에러 및 분산 정보를 이용해 결정한다. 구성되지 않은 블록들은 작은 움직임을 고려해 탐색 영역을 ± 2 까지 설정한다.
6. 움직임 추정 : 각 영역별로 구성된 가변 탐색 영역을 이용해 8x8 블록별로 설정된 탐색 영역 범위 내에서 움직임 추정을 수행한다. 이 과정에서 단계 4에서 탐색한 블록은 재탐색하지 않는다.
7. 확장 재추정 : 추정된 블록 중 에러가 임계치(Th2)보다 크면 ± 16 까지 확장하여 재추정한다.

if (MadError(B_i) > Th2)

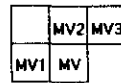
ExpandEstimate();

확장 추정 탐색 범위는 그림 1에 예시되어 있다.

III. 영역 재병합 및 움직임 벡터 부호화

이 장에서는 분할된 영역 정보를 이용해 B-Picture의 움직임 정보를 부호화 하는 단계로 영역 재병합 과정과 최종의 마커 추출 및 영역의 위치 정보 추출 후 허프만(Huffman)기법을 이용해 움직임 정보를 부호화 하게된

1. 움직임 정보를 이용한 병합 : 병합되지 않은 블록 중 제로가 아닌 벡터를 갖는 블록은 전에 구성된 영역과 8방향으로 인접하면 병합하고 구성된 블록의 내부에 속한 블록들도 움직임을 갖는 영역의 일부분으로 간주하여 병합하여 마커를 갱신시킨다. 이 단계에서도 제로 벡터가 아닌 작은 고립 영역들은 II장의 단계 3에서처럼 병합되지 아니하고 제거된다.
2. 구성된 영역의 마커 및 위치 정보 추출 : 재병합된 영역들이 서로 8방향으로 연결되면 다시 재병합 과정을 거친다. 마지막으로 마커를 갱신시키고 각 영역별로 영역을 포함하는 최소의 사각형 위치 정보를 구한다.
3. 움직임 벡터 부호화 : 움직임 벡터의 부호화는 허프만 부호화(Huffman Coding)를 통해 통계적인 압축을 수행하는 것이 일반적이는데, 본 연구에서는 H.263에서 적용하는 예측 코딩(Differential coding) 기법을 이용한다.



MV Current Motion Vector
MV1 Previous Motion Vector
MV2 Above Motion Vector
MV3 Above Right Motion Vector

그림 2. 움직임 벡터 예측

$$Px = \text{Median}(MV1x, MV2x, MV3x)$$

$$Py = \text{Median}(MV1y, MV2y, MV3y)$$

$$MVDx = MVx - Px;$$

$$MVDy = MVy - Py;$$

4. B-Picture의 움직임 부호화 : 분할된 영역별로 위치 정보 및 움직임 벡터를 부호화 한다. 위치정보는 영역을 포함하는 사각형 정보로서 CIF 이미지의 경우는 각 위치당 6bit만을 할당하면 총 2BYTE면 표현 가능하다. 신택스 구조는 아래와 같다.

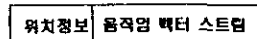


그림 3. 움직임 벡터 신택스 구조

H.263 기반의 B-Picture에 비해 비트율이 적어지며, 구성된 영역의 수가 적고 영역의 크기가 적을수록 비트 효율이 커진다.

5. 모드 결정 : 손실이 심한 영상이나 전체적으로 움직임 영역이 복잡한 영상은 영역 분할 정보를 이용할 경우 움직임 벡터 정보가 기존의 방법보다 커지

는 경우가 있다. 이 경우를 위해 기존 방법과 제안한 알고리즘간의 모드 결정 단계를 두어 효율을 얻는다.

6. 에러 이미지 보상 부호화 : 움직임 추정 후 생기는 에러 이미지는 기존의 H.26p나 MPEG1,2,4에서 처럼 DCT기법을 이용해 처리를 한다. 본 논문에서는 분할된 영역들에 속하는 블록들에 대해서만 에러 보상을 수행하는데 에러가 심한 블록만을 보상한다. 기존의 방식과는 다르게 매크로 블록별로 수행하지 않고, 다음 장에서 설명할 가변 영역 분해능 부호화(Scalable Region Coding)를 지원하기 위해서 영역별로 부호화 된다.

IV. 가변 영역 분해능을 위한 의미 영역 인식기법

화상회로나 화상전화 같은 영상은 크게 전경(Foreground image)과 배경(Background Image)으로 구성되는데 전경은 배경에 비해 의미가 크다고 볼 수 있다. 또한 전경을 얼굴과 몸통 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 주관적인 화질을 유지할 수 있다. 실제로 위에서 제시한 영역 분할 결과를 살펴보면 구성된 영역들은 얼굴 영역과 손을 중심으로 한 신체영역으로 분할된다.

기존의 H.263이나 MPEG-4에서는 다양한 시공간 가변 분해능(Spatial and Temporal Scalability)을 지원하는데, 이 경우 다소 전체영상의 화질 저하를 초래하는 단점이 있다. 그러므로, 분할된 영역중 의미가 큰 얼굴과 손을 중심으로 한 영역들에 우선권을 부여해 개별적인 부호화 및 전송을 위한 인식 기법을 제시한다. 인식은 살색 정보를 이용한다.[5] 세부적인 알고리즘은 아래와 같다.

1. 색 신호 성분 U, V를 이용한 살색 영역 인식 : 사람의 피부색은 인종마다 다소의 차이가 있지만 색 신호 좌표 계에서 보면 일정한 영역에 분포하게 된다.[5] 이 영역을 이용해 분할된 영역들에 대해서 블록 내부에 속하는 화소들을 살색화소로 인식한다. 살색 좌표 계는 아래 그림과 같다.

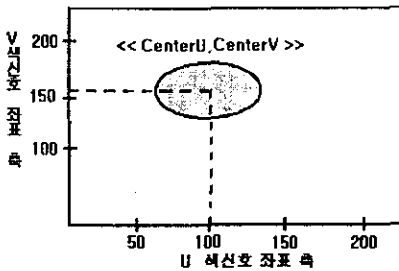


그림 4. 살색 좌표계

그림 4에서의 타원 영역은 색 신호가 존재하는 영역으로 살색화소 결정 여부는 아래 식을 만족하면 살색 화소로 간주한다.

$$\frac{(U - CentreU)^2}{RadU^2} + \frac{(V - CentreV)^2}{RadV^2} < 1$$

여기서, U는 화소의 U값, V는 화소의 V값

CentreU, CentreV는 타원의 중심값

RadU, RadV는 타원의 축 반지름

다음으로, 살색 블록 결정 여부는 블록 내부의 화소수가 미리 정한 임계치(Th3) 이상일 경우에 살색 블록으로 결정한다.

2. 가로축 창(Window)설정을 통한 대략적인 얼굴 영역의 분석 : 얼굴 영역은 일반적으로 화면 중앙에 위치하므로 가로축으로 창을 설정해 단계 1에서 찾은 살색 영역 중 창 내부에 속한 분할 영역(Segmented Region)들을 결정한다.
3. 세로축 창 설정을 통한 얼굴영역 결정 : 위 단계들을 통해서 판정된 살색 영역들이 얼굴이 아닌 손 같은 다른 영역일 수 있으므로 세로축으로 창을 설정해 얼굴 영역을 결정한다.
4. 이렇게 결정된 얼굴 영역에 대해서는 최 우선적으로 부호화 하고 전송 시 우선권을 주고 나머지 손과 같은 살색 영역에 대해서는 배경과 같은 의미가 덜한 영역에 비해 우선권을 부여한다.

V. 실험 결과 및 검토

제안한 알고리즘의 성능 평가를 위해 352x288 Salesman(30Frame), 360x288 Claire(80Frame), 360x288 Miss America(40 Frame) 실험 이미지를 사용하였다. 그림 5에 실제 실험에 사용된 영상이 예시되어 있다.



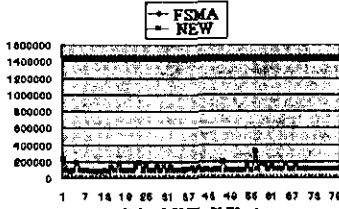
그림 5. 실험 영상

제안한 알고리즘 중 분할 기반 움직임 추정의 결과 측정은 매칭되는 블록 수와 PSNR로 표1에 나타내고 가변 영역 결정 시 이용한 영역 분할 영상 및 결과는 Claire 이미지에 대해 그림 6에 예시되어 있다. 그림 6(b)에서 볼 수 있듯이 Claire 영상의 경우 제안한 알고리즘이 전역 탐색 기법에 비해 평균적으로 블록 매칭 수가 10.3배 작게 나타났고, PSNR은 제안한 알고리즘이 0.4dB 정도 낮게 나타났다. 전체 결과는 표 1에서 볼 수

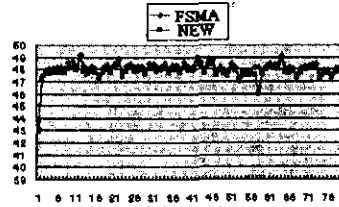
있듯이 매칭된 블록 수는 제안한 알고리즘이 상당히 작게 나타났고 PSNR면에서는 0.2~0.6dB 정도 낮게 나타났다.



(a) 본촬영 영상



(b) 매칭된 블록 수



(c) PSNR

그림 6. 움직임 추정 결과(Claire)



(a) 본촬영 영상

(b) 인식된 의미 영역

그림 7. 살색 영역 인식

다음으로 제안한 알고리즘인 B-Picture 움직임 부호화 결과는 표2에 예시되어 있다. 표2에서 볼 수 있듯이 제안한 알고리즘이 기존의 방법에 비해 비트량이 적게 나타났다. Salesman 영상의 경우에는 프레임 중간에 손상된 실험 영상에 의한 구성된 영역이 복잡해져 이 부분에서 비트량이 두배로 발생해 그다지 압축이 크게 되지는 않았지만 3장의 단계 5에서 제시한 모드 결정을 통해서 단점을 극복 할 수가 있었다. 마지막으로 의미를 갖는 얼굴 및 손 영역 인식의 결과영상은 그림 7에 예시되어 있다. 모의 실험은 분할 영역 중 얼굴 영역을 먼저 인식하고 얼굴 영역에 대해서 최 우선권을 부여하고 다음으로 결정된 손 영역에 대해서는 다음 우선권을 부여했고 이를 이용해서 선택적으로 부호화를 했다.

표 1. 평균 블록 매칭수 및 PSNR

시험 영상	평균 블록 매칭 수		평균 PSNR	
	FSMA	NEW	FSMA	NEW
Claire	1448272	111029	47.77	47.52
Miss America	1448272	111746	44.25	43.65
Salesman	1415040	302452	43.01	42.94

표 2. 움직임 정보 부호화 비트량

시험 영상	평균 부호화 비트량(BYTE)	
	기존 방법	제안한 방법
Claire	1209.9	701.48
Miss America	2032	484
Salesman	990	883.31

VI. 결론

본 논문의 알고리즘은 대역폭이 매우 제한된 이동 통신 망의 초 저속 영상 부호화(Very Low Bit rate Video Coding)에 적용하기 위한 목적으로 제안되었다. 움직임 추정은 화상 전화와 같은 영상의 움직임은 상당히 적다는 사실에 기반을 두어 영역 분할 정보를 이용해 고속으로 처리했고 결과는 FSMA에 비해 상당히 빠르며 PSNR은 다소 낮아졌다. 다음으로, B-Picture 움직임 정보 부호화 알고리즘에 있어서도 분할된 영역 정보를 이용하여 부호화 하는 기법이 기존의 기법보다 비트율이 약 1/2에 정도 적게 나타났다. 하지만 손떨림이나 손실이 있는 영상의 경우는 영역 분할이 복잡하게 이루어져 발생 비트율이 커지는 단점이 있으므로 이 부분의 보완이 필요하다. 마지막으로, 얼굴이나 손 같은 의미 영역의 인식은 임계치의 대략적인 설정 하에서도 제대로 이루어 졌지만 앞으로의 보완 점은 영역별 가변 분해능(Region Scalability)을 위해 일정한 비트량 생성을 위한 Rate-Distortion 측면에서의 접근이 필요하겠다.

참고 문헌

1. K.R RAO, J.J HWANG "Techniques and Standards for Image, Video, and Audio Coding", Prentice Hall 1996.
2. Y. Baek, H.-S.Oh, and H.K Lee "Block Matching Criterion for Efficient VLSI Implementation", Electronic Letters, 32(13), 1184-1185 (jun.1996)
3. B. Liu and A.Zaccarin, "New Fast Algorithms for the Estimation of Block Motion Vectors" IEEE Trans. On Circuits and Systems for Video Technology 3(2), 148-157, 9Apr.1993)
4. Milan Sonka, Vaclav Hlavac, Roger Boyle, "Image Processing, Analysis and Machine Vision". Champman & Hall Computing
5. Demas Sanger, Takuya Asada, Hiedeki Haneishi and Yoichi Muyaake, "Facial Pattern Detection and Its Preferred Color Reproduction", The 2nd IS&T/Sid Color Imaging conference, pp. 149-152, 1994