

객체 기반 스테레오 동영상 부호화

권순익, 배태민, 한규필, 하영호
 경북대학교 전자전기공학부
 702-701 대구광역시 북구 산격동 1370번지
 e-mail: yhha@ee.kyungpook.ac.kr

Object-based Stereo Sequence Coding

Sun-Kyu Kwon, Tae-Min Bae, Kyu-Phil Han, and Yeong-Ho Ha
 School of Electronic & Electrical Eng., Kyungpook Nat'l Univ.
 1370 Sankyuk-dong, Puk-gu, Taegu 702-701, Korea

요약

본 논문에서는 객체기반의 스테레오 동영상 부호화기를 제안한다. 제안된 부호화기는 움직임, 변이, 그리고 움직임 정보를 이용하여 객체를 추출하고 객체의 움직임과 변이 정보를 예측하여 부호화함으로써 기존의 블록기반 방법에 비해 정보량을 줄일 수 있었다. 또한 새롭게 발생하는 영역과 기존 영역의 소멸 및 변형은 부호화 효율을 떨어뜨리게 되는데, 이를 해결하기 위해 새로운 객체생성 방법이 제안되었다. 실험을 통하여 제안된 객체기반 부호화기는 기존의 블록기반 부호화기에서 나타나는 부호화 및 모스키노 현상을 제거할 수 있었고 복원 화질과 부호화율을 개선시킬 수 있었다.

I. 서론

정보화가 가속화됨에 따라 문자뿐만 아니라 오디오 및 비디오를 포함하는 멀티미디어 정보에 대한 수요가 급증하게 되었고 이런 복합정보의 효율적인 저장과 전송을 위해 활발한 연구가 진행되고 있다. 특히 실시간 영상통신(visual communication)에 관계된 JPEG, MPEG 및 MPEG2 그리고 H.263에 기반한 2-D 디지털 영상통신 기술들이 급속한 발전을 거듭하여 현재 여러 제품에 상용화되고 있다. 그러나 이 기술들은 가상현실의 도래로 인해 수반된 3차원적 현실감을 다루기에는 불충분하다고 할 수 있다. 현재 정보통신의 발전 방향을 볼 때, 향후 가상현실과 입체영상에 대한 사용자의 수요증가로 인해 스테레오 시각에 기초한 3차원 영상압축기술과 3차원 영상표현시스템이 연구된 것이며 이들이 미래 영상매체의 핵심기술이 될 전망이다.

가상 현실에 대한 관심이 증가함에 따라 3차원 영상을 경험할 수 있는 입체 영상 부호화기가 필요하며 이를 스테레오 동영상은 영상면간의 시간적 중복성과 두 스테레오 영상사이의 공간적 중복성을 가지고 있기 때문에 정확한 움직임과 변이를 추정한다면 전송효율을 극대화시킬 수 있다.

스테레오 동영상영역은 좌우 두 개의 단일 동영상영역으로 이루어져 있으므로 정보량이 기존의 동영상들 보다 2배 이상 크므로 효율적인 저장이나 전송을 위해서 좌우 영상사이의 공간적 중복성과 각 영상면 내에서의 시간적 중복

성을 효과적으로 제거해야만 한다. 초기에 연구된 스테레오 영상 부호화기는 기존의 MPEG방식을 스테레오 영상에 적용하여 블록 정합 방식의 시간적 중복성인 움직임(motion)과 공간적 중복성인 변이(disparity)를 각각 예측함으로써 압축하게 된다. 그러나 변이 정보는 움직임 정보와 특성이 달라 블록 정합 방식으로는 정확한 변이를 추출하기 어렵고, MPEG방식의 부호화는 시공간적인 상관성을 고려하기가 어렵기 때문에 압축률과 예측능력에 한계가 있다. 또한, 변이 정보는 영상의 3차원 정보를 가지고 있기 때문에 효과적으로 변이를 예측하여 부호화할 수 없다. 그래서 점차 스테레오 영상의 특성을 고려할 수 있는 객체(object) 기반의 부호화 기법에 대한 연구가 증가하고 있다.^[1]

객체기반 부호화 기법은 움직임과 변이가 동일한 부분을 영상에서 객체로 추출하여 객체 단위로 부호화 함으로써 영상을 압축하는 방식이다. 객체기반 부호화는 MPEG 방식의 2차원적 움직임 예측 부호화에 비해 3차원적인 움직임을 예측하기 때문에 영상을 정확하게 예측할 수 있다. 또한 지금까지 연구되어진 스테레오 정합에 대한 종래의 기법을 쉽게 적용할 수 있으므로 정확한 변이 예측과 추출된 변이 정보의 압축에서도 효율적인 부호화를 할 수 있다.

본 논문에서는 움직임이나 변이가 제대로 추정되지 않는 부분을 전체처리정성을 통하여 찾고 객체추출에서 이 부분을 제외시킴으로써 잘못된 객체추출에 의한 영상에추오차를 줄일 수 있었다. 또한 움직임과 변이의 유사성과 함께 밝기값을 고려하여 좀 더 정확한 객체를 추출하였다. 객체기반 부호화의 경우 새로운 영상영역의 발생과 폐색영역(occlude region), 객체에추오차가 큰 영역에 대한 객체생성이 요구되는데, 본 논문에서는 이런 부분에 대해서 객체추출과정을 수행하여 새로운 객체를 추출하였다. 부호화 과정은 그림 1에서와 같이 먼저 입력된 좌우영상에 대해 전처리(preprocessing)부에서 객체추출이 어려운 영역을 찾고 변이와 움직임을 화소단위로 예측한다. 예측된 변이와 움직임이 유사한 영역을 객체로 추출하고 객체단위의 움직임과 변이를 추정한다. 우영상은 객체의 움직임 정보를 이용하여 영상에추오하고 폐색영역과 예측오차가 큰 영역에 대해서 객체생성을 수행한다. 좌영상에서는 변이정보와 우영상의 객체 움직임 정보를 이용하여 좌영상의 움직임 정보를 유도하여 영상을 예측한다. 또 주목되는 전체영역오차를 줄이기 위해서 객체를 다시 구한다.

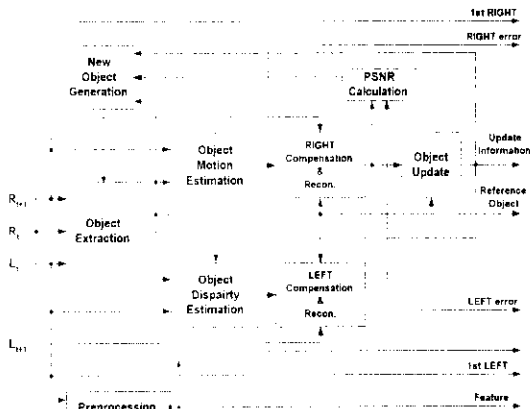


그림 1. 재안된 객체 기반 스테레오 동영상 분호화기의 블록도

II. 객체 추출을 위한 전처리과정

객체를 추출할 때 움직임 정보나 변이 정보를 이용하여 밝기값이 유사하면서 움직임이 거의 없는 비객체영역에서는 정확한 객체를 추출하기 어렵다. 비객체영역에서의 잘못된 객체추출은 영상내의 객체를 증가시키며 움직임 및 진동정보량을 증가시킨다. 그러므로 비객체영역을 움직임이나 변이 추정 여자에 추출해냄으로써 객체 추출에서 제외시킬 수 있다.^[1]

움직임이나 변이 정보 예측에서 예측된 값의 신뢰도를 높이기 위해서는 예측된 블록의 밝기 차이가 작을수록 좋다. 이러한 가정에 기반하여 탐색영역에서 추정되는 블록의 밝기 차이에 대한 편차를 구함으로써 객체영역과 비객체영역을 구분할 수 있다.

$$\text{Var} \left[\sum_{(k,l) \in A} (R(i,j) - L(i+k, j+l)) \right] < T_{thr} \quad (1)$$

여기에서 A 는 탐색영역을, T_{thr} 는 문턱치값을 의미한다.

블록 블록 밝기값 차이의 편차에 대한 고려만으로는 정확한 비객체영역의 분리가 불가능하다. 그러므로 이런 일의으로 분류된 후 밝기 평균값이 유사한 주변블록을 병합하는 과정을 추가하였다.

$$\frac{1}{N} \sum R(i,j) - \frac{1}{N} \sum_{(k,l) \in A} R(k,l) < T_{thr} \quad (2)$$

은 병합에 사용된 수식이다. 여기서 N 은 한 블록 내의 화소의 개수, $A(i,j)$ 는 밝기차 편차에 의해서 예측된 비객체영 (i,j)의 인접영역, 그리고 T_{thr} 는 문턱치값이다.

III. 객체 추출

객체기반 분호화기는 객체단위로 움직임과 변이 정보를 추정해서 전송하기 때문에 객체추출 성능이 중요하다.

다. 움직임 정보와 변이 정보가 유사한 영역을 객체로 추출하면 객체단위의 움직임 및 변이 정보추정의 신뢰도(reliability)를 높일 수 있다. 그림 2는 객체추출을 위한 순서도이다.

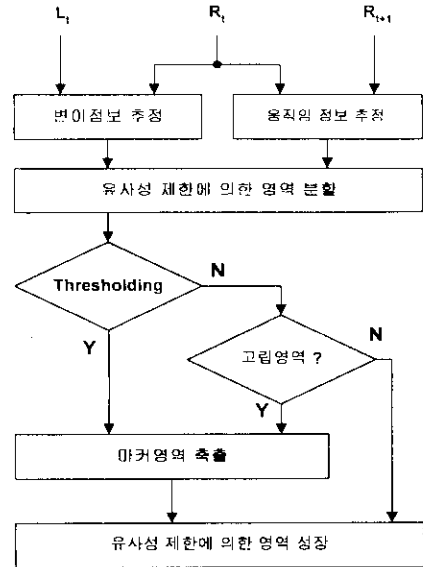


그림 2. 객체추출과정의 순서도

1. 변이 정보 예측

정확한 객체추출을 위해서는 객체의 경계부분이 명확히 구분되어야 하기 때문에 변이 정보를 화소단위로 추정해야 한다. 또한 탐색범위의 제한없이 변이 정보를 추정하게 되면 연산량이 증가하고 정확한 추정이 어려워진다. 그러므로 적절한 탐색범위를 구하기 위해서 계층적 방법을 사용하였다.

객체영역내의 화소들에 대한 변이정보 추정은 평탄제약(smoothness constraint)을 적용함으로써 정확한 변이 정보를 얻을 수 있다. 그러므로 아래의 비용함수를 최소화 하는 변이값을 선택하였다.

$$J = \sum_i \sum_j (R(i,j) - L(i,j+d)) + \alpha |d_b - d_l| + \beta |d_l - d_r| \quad (3)$$

여기서 d_b 는 이진 화소의 변이값, d_l 은 이진 라인의 변이값을 나타내고 d 의 범위는 $[d_{min}, d_{max}]$ 이다.^{[1][2]}

2. 움직임 정보 예측

블록의 움직임이 없고 카메라가 움직이는 경우, 블록 검색 방법으로 움직임을 예측하면 전영역에서 움직임이 동일하다. 이런 경우에는 예측된 움직임정보로 객체를 추출할 수 없다. 그러나 광류(optical flow)를 사용하면 밝기값의 변화에 대한 움직임이 추정되기 때문에 위와 같은 경우에서도 전영상에서의 광류는 다르다. 즉 광류는 전영상에 관계없이 객체추출이 가능하다.^[3]

3. 객체분할

본 문헌에서는 변이와 움직임이 유사한 영역을 객체로 정의하였고 객체분할에서는 예측된 변이와 움직임뿐

만 아니라 밝기정보를 고려하였다. 객체분할은 먼저 변이와 움직임이 동일한 영역에 대해서 단일큐(queue)를 사용하여 영역을 분할하지만 분할된 영역들이 모두 객체가 될 수 있는 것이 아니라, 영역의 크기가 문턱치보다 클 때 객체추출을 위한 마커가 된다. 그러나 크기가 문턱치보다 작아도 그 영역 내에서 변이와 움직임 정보가 주위의 값들과 차이가 크다면 이 영역은 독립된 영역으로서 마커가 된다. 추출된 마커를 정정시키기 위해서는 다중큐(multi-queue)를 사용하는데, 영역성장시 성장 정보는 변이와 움직임정보에 밝기정보가 첨가되어야 한다. 왜냐하면 밝기 차이가 큰 경계부분은 서로 다른 객체의 경계가 될 가능성이 크기 때문이다. 또한 움직임에 의해 사라질 확률이 큰 수평수직 방향의 좁은 선 영역은 미리 중간값필터(median filtering)로 제거시켰다.

IV. 객체기반 영상 예측

1. 좌우영상복원

첫 번째 프레임의 좌우영상 그리고 두 번째 프레임의 우영상으로부터 추출된 객체를 기준(reference)객체로 정의한다. 이 객체의 기준영상은 첫 번째 프레임의 우영상이 되고 다음 프레임 우영상은 객체단위의 움직임추정함으로써 복원할 수 있다. 그러나 좌영상의 경우 기준객체의 기준영상이 우영상이므로 좌영상에 대한 객체단위의 움직임정보를 추정할 수 없다. 기준객체를 가지고 객체단위로 변이정보를 추정하였다 하더라도 변이정보와 움직임정보를 비교하여 화질이 떨어지게 되므로 영상예측에 직접 적용할 수 없다. 좌영상에 대한 움직임정보는 아래의 식에서의 값이 현재 프레임에서의 변이정보, 다음 프레임에서의 변이정보와 우영상에 대한 움직임정보를 이용하여 구할 수 있다.

$$D_t = R_t - L_t \quad (4)$$

$$D_{t+1} = R_{t+1} - L_{t+1} \quad (5)$$

$$M_v = R_{t+1} - R_t \quad (6)$$

$$M_l = L_{t+1} - L_t = D_{t+1} - D_t + M_v \quad (7)$$

여기서 D_t 는 현재 프레임의 변이, D_{t+1} 은 다음 프레임의 변이, M_v 은 우영상의 움직임, M_l 은 좌영상의 움직임, L_t 은 좌영상의 밝기, R_t 은 우영상의 밝기를 나타낸다.

좌우영상을 복원하기 위해서는 첫 번째 프레임의 좌우영상은 영상내부부호화(intraframe coding)해서 전송해야 하고 기준객체에 대한 정보, 객체단위의 움직임정보, 변이정보, 및 움직임예측오차를 전송해야 한다.

2. 객체의 갱신

여러 프레임용 시다면서 기준객체는 객체단위 움직임정보에 의해서 변화하게 된다. 객체끼리 겹쳐지기도 하고 객체가 존재하지 않는 영역도 생기게 된다. 또한 하나의 객체 내에서 변이와 움직임이 상이해서 두 개 이상의 객체로 분할되어야 할 필요가 있다. 객체가 존재하지 않는 부분이 커지게 되면 그 부분은 인트라부호화해야 하기 때문에 전송정보량은 급격히 증가하게 된다. 이 부분에 대해서는 화소단위의 변이와 움직임정보가 동일한 영역을 추출해서 새로운 객체로 인식하였다. 하나의 객체 내에서 움직임이 동일하지 않은 객체는 각 객체별로 PSNR을 구함으로써 찾아낼 수 있다. 이 객체에 대해서는 이런 객체추출 방법과 같이 변이정보와 움직임정보

고리고 밝기값 정보를 이용해서 다시 객체를 분할하였다. 위의 객체갱신을 거친 후에도 객체의 움직임에 의한 오차는 프레임이 지날수록 계속 누적되고 특히 움직임이 갑자기 커지는 프레임에서는 객체갱신의 의미가 없게 된다. 이것은 부호화후에서 한 프레임마다 복원된 좌우영상의 PSNR을 구해서 문턱값보다 작으면 기준객체를 다시 구해서 전송함으로써 해결된다. 객체갱신과 동시에 추가될 정보량은 새로운 객체에 대한 정보와 분할된 객체에 대한 정보이다.

IV. 실험 및 결과

384×384 크기를 가지는 25프레임의 'man' 과, 640×512 크기의 25프레임의 'aqua' 영상이 실험에 사용되었다. 실험에 사용된 스테레오 영상은 극상선(epipolar line) 제한 조건을 만족시키므로 일차원 탐색만을 수행하여 변이 예측을 하였다. 그림 3의 (d)에서 'man' 영상의 기준객체 개수는 17개이고 객체갱신에 의해 그 개수는 20, 23, 31로 증가한다. 단 이 경우 객체크기 제한은 50 화소이다.

'man' 영상은 변이가 -1(x)에서 -40의 범위 내에 있을 때만 움직임은 범위가 훨씬 좁다. 그렇게 되면 움직임정보로 복원된 우영상이 변이정보와 우영상의 움직임정보로 복원된 좌영상보다 더 좋은 화질을 나타낸다. 표에서도 그 사실을 확인할 수 있다. 'aqua' 영상의 경우 변이가 -15에서 0의 범위 내에 있어서 복원된 좌영상의 화질이 우영상에 비해 떨어지지 않는다. 그림 4와 5로부터 객체기반 부호화로 복원한 영상에서는 블록기반 부호화에서 나타나는 블록화현상이나 모스기호현상이 나타나지 않고 정확한 경계부분을 복원한다.

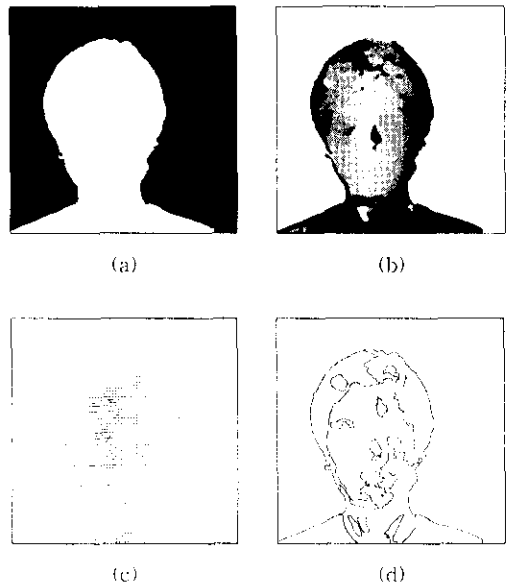


그림 3. 'man' 스테레오 영상. (a) 전처리과정의 결과 영상 (b) 객체추출을 위해 예측된 변이도 (c) 예측된 광류 (d) 추출된 객체도

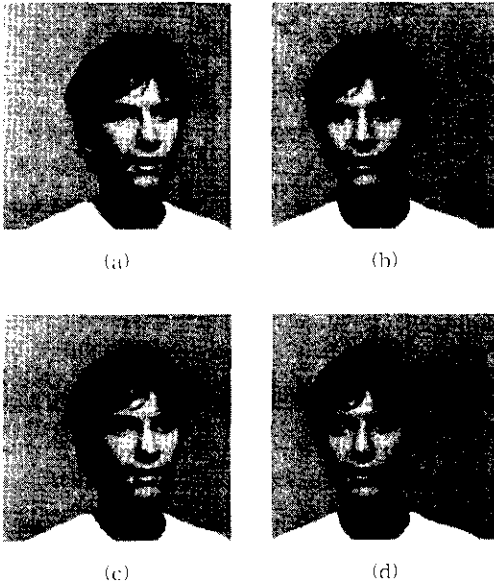


그림 4. 'man' 스테레오 영상. (a) 원래의 두 번째 프레임 좌영상 (b) 원래의 두 번째 프레임 우영상 (c) 예측된 두 번째 프레임 좌영상 (d) 예측된 두 번째 프레임 우영상

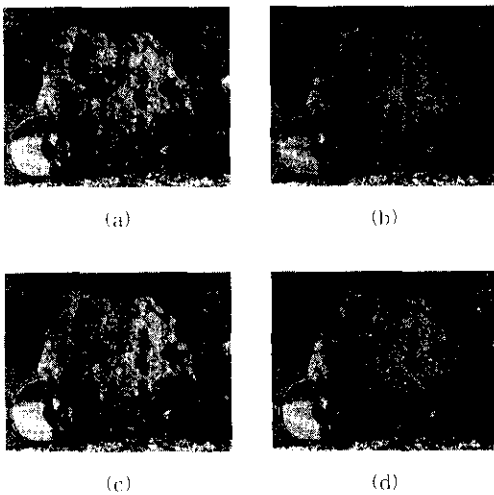


그림 5. 'aqua' 스테레오 영상. (a) 원래의 두 번째 프레임 좌영상 (b) 원래의 두 번째 프레임 우영상 (c) 예측된 두 번째 프레임 좌영상 (d) 예측된 두 번째 프레임 우영상

표 1. 'man' 스테레오 영상에 대한 PSNR.

	LEFT	RIGHT
2nd frame	41.359759	43.120504
3rd frame	38.588379	43.359591
4th frame	36.089604	40.349291

표 2. 'aqua' 스테레오 영상에 대한 PSNR.

	LEFT	RIGHT
2nd frame	42.110204	39.679823
3rd frame	37.338991	37.716877
4th frame	36.991370	36.991370

V. 결론

본 논문에서는 스테레오 동영상 부호화기를 제안하였다. 제안된 부호화기는 움직임, 변이, 그리고 밝기 정보를 이용하여 객체를 추출하고 객체의 움직임과 변이 정보를 예측하여 부호화함으로써 객체기반 부호화기를 구현했다. 부호화시 새로운 객체생성 방법을 제안함으로써 새롭게 발생하는 영역과 기존 영역의 소멸 및 변형에 의한 부호화 효율저하를 줄일 수 있었다. 또한 전처리과정을 통하여 움직임 및 변이 정보 예측이 어려운 영역을 객체추출과정에서 제거하여 영상에중요자를 줄일 수 있었다. 제안된 방법의 실험 결과는 객체기반 부호화기가 기존의 블록기반 부호화기에서 나타나는 블록화 및 모스키노 현상을 제거할 수 있고 복원화질과 부호화율을 개선시킬 수 있다는 것을 보여주고있다.

참고문헌

- [1] A. Puri, R. V. Kollarits and B. G. Haskell, "Compression of stereoscopic video using MPEG-2," *Critical Reviews conference on Standards and Common Interfaces for Video Information Systems*, published by SPIE, Oct. 1995.
- [2] M. Ebroul Izquierdo, "Stereo image analysis for multi-viewpoint telepresence applications," *Signal Processing Image Communication*, vol. 11, pp. 231-254, 1998.
- [3] Stathis Panis, Manfred Ziegler, and John P. Cosmas, "The use of stereo and motion in a generic object-based coder," *Signal Processing Image Communication*, vol. 9, pp. 221-238, 1997.
- [4] Dimitrios Tzovaras, Nicos Grammalidis, and Michael G. Strinziis, "Disparity field and depth map coding for multiview 3D image generation," *Signal Processing Image Communication*, vol. 11, pp. 205-230, 1998.
- [5] Berthold K.P. Horn and Brian G. Schunck, "Determining Optical Flow," *Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*, vol. 17, pp. 185-203, 1981