

동영상에서의 내용기반 메쉬를 이용한 모션 예측

김 형 진, 이 동 규, 이 두 수
한양대학교 전자 공학과
전화 : (02) 2290-0358 / 팩스 : (02) 2298-1796

Content Based Mesh Motion Estimation in Moving Pictures

Hyung-Jin Kim, Dong-Gyu Lee, Doo-Soo Lee
Dept. of Electronic engineering, Hanyang University
E-mail : f5tiger@ses.hanyang.ac.kr

Abstract

The method of Content-based Triangular Mesh Image representation in moving pictures makes better performance in prediction error ratio and visual efficiency than that of classical block matching. Specially if background and objects can be separated from image, the objects are designed by Irregular mesh. In this case this irregular mesh design has an advantage of increasing video coding efficiency.

This paper presents the techniques of mesh generation, motion estimation using these mesh, uses image warping transform such as Affine transform for image reconstruction, and evaluates the content based mesh design through computer simulation.

keyword :

Content-based, Triangular mesh, Irregular mesh, affine transform, image reconstruction

I. 서론

일반적으로 동영상은 통신 또는 매체를 통해서 원 영상을 전송하거나 전달하기에는 신호 특성상 많은 데이터를 가지고 있다. 따라서 동영상 데이터를 전송

하기 위해서는 효율적인 압축 코딩이 필수적이다.

동영상의 시간, 공간적으로 생기는 중복 (Redundancy) 성분을 제거함으로써 코딩효율을 높일 수 있다.

이에 현재 가장 널리 사용되어지는 H.261, H.263 표준안에서는 블록 매칭(Block Matching)을 통한 시간상 압축 코딩을 이용한다. 하지만 이러한 이동 블록 (Translational Block) 방식은 모션 예측(Estimation)과 보상(Compensation)에서의 구조적 간결성으로 인해 많이 사용되지만 영상자체의 회전(Rotation)과 비례 확대축소(Scaling)에 대처 할 수 없다. 또한 블록 상호간에 연관성을 가지지 못함으로 블록의 구조적인 단점으로 인하여 블록 왜곡현상(Blocking Artifact)을 수반한다. 그러므로 블록 매칭을 이용한 압축코딩은 화면의 질이 저하되는 문제를 가지고 있다.

이 같은 단점을 극복하기 위한 방법으로 영상 내용에 근거한 삼각 메쉬(Triangular Mesh)방법이 제안되었다. 이를 이용한 영상 예측과 복원된 기존의 블록 왜곡현상을 줄이고 영상 예러 울(Peak Signal to Noise Ratio)을 높일 수 있다.

본 논문에서는 Content base Mesh를 구성하기 위한 영상의 노드의 위치를 결정하는 방법과 삼각 메쉬 (Triangular Mesh) 모델을 이용한 영상 모션 예측과 보상에 관한 내용을 다룬다.

II. Mesh 생성을 위한 노드 생성

동영상에서의 Mesh 를 생성하는 방법으로는 메쉬의 간격을 일정하게 생성시키는 Regular Mesh 생성 방법과 내용기반에 따른 일정하지 않은 크기의 메쉬로 생성하는 Irregular Mesh 방법이 있다. 일반적으로는 메쉬의 간격을 일정하게 설정하는 Regular 메쉬 방법이 사용되나 불필요한 메쉬 성분도 갖게되는 단점이 있다. 한편 Irregular 메쉬 방법은 생성 과정에서 연산량은 많지만 내용 성분(Content based)에 근거하여 메쉬 생성을 하면 불필요한 메쉬 성분을 줄일 수 있으므로 낮은 비트 율(Bit Rate) 에서도 영상의 내용을 전달할 수 있다.

이러한 Irregular 메쉬를 생성하기 위한 첫 단계는 내용기반 노드를 생성하는 일이다. 어떻게 이러한 노드를 효율적으로 생성하는가에 따라 메쉬 기반 모션 예측과 보상의 효율이 결정된다.

내용 기반 노드를 잡기 위한 단계는 다음과 같다.

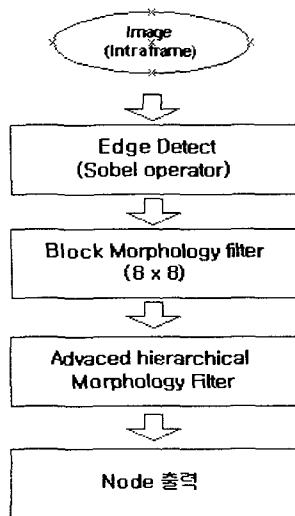


그림1. 노드를 생성 하기 위한 단계

이미지의 공간적 에지 (Spatial Edge) 를 $SE(x,y)$ 라면 초기노드의 할당은 식(1) 에 의해서 결정된다.

$$SE(x,y) = \sqrt{\left(\frac{dI(x,y)}{dx}\right)^2 + \left(\frac{dI(x,y)}{dy}\right)^2} < T_{SE=0} \quad (1)$$

8x8 블록 단위로 노드를 할당하기 위해 블록 단위로 Morphology 필터로 Thinning을 하고 그 중에 중심과 가장 가까운 노드를 두 번째 노드로 할당을 한다. 블록 단위로 생성된 노드에서 이 블록을 3x3, 8x8 블록으로 다시 정의하고 이 영역의 노드가 존재할 경우

각 노드의 위치 상관 관계에 따라 노드를 하나씩 소개해 간다(Advanced hierarchical morphology filtering).

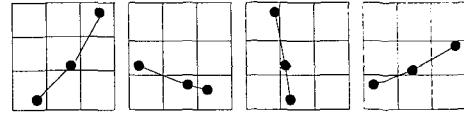


그림2 가운데 점을 없앨 수 있는 경우

그림2에서 보듯이 존재하는 각 노드가 중심을 기준으로 각도 20° 안에 들어 갈 때 가운데 노드를 없앤다. 3x3 블록이 끝나면 8x8 블록으로 행한다.

Advanced hierarchical morphology filtering을 통해서 굴곡진 곳에서는 많은 노드를 할당하고 직선성분의 노드를 줄일 수 있다. 이렇게 수행되어진 노드들은 물체와 같은 영역에서는 조밀하게 노드를 생성하고 단조로운 배경과 같은 영역의 노드는 상대적으로 노드가 넓게 생성 된다.

III. Mesh model

motion estimation & compensation

1. Triangulation

Mesh 단위로 움직임 예측과 보상을 하기 위해서는 선택된 노드로부터 Mesh를 생성(Triangulation) 시키고 이에 따른 예측과 복구 과정을 거쳐야 한다.

가장 많이 사용되는 방법으로 Delaunay triangulation [9] 이 쓰이지만 이는 임의로 노드를 추가하는 경우가 발생하므로 가장 가까운 노드를 서로 연결해 가는 Triangulation 과정을 거쳐서 Mesh를 생성한다.

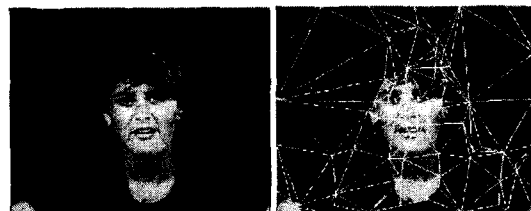


그림3. 원 영상(좌) Triangulation 한 그림(우)

그림 3에서 보듯이 배경 부분은 넓은 삼각 메쉬로 구성되고 물체(사람)의 경우는 세밀한 삼각 메쉬로 구성되어 진다.

2. Triangular motion estimation & compensation

움직임을 표현하는 모델은 움직임을 정의하는 방식에 따라 크게 매개변수 모델(Parametric model) 과 비 매개변수 모델 (Non parametric model)로 나뉠 수 있다. 이 중에서 삼각 메쉬를 기반으로 움직임을 정의 할 경우 매개변수 모델 방식인 Affine 모델[3]을 사용한다. Affine 모델은 6개의 변수를 가지고 모션을 표현하는 모델로서 이동(Translation) 과 회전 (Rotation)을 표현 하는데 적합한 모델이다. Affine 모델은 다음과 같다.

$$x = a_1x' + a_2y' + a_3 \quad (2)$$

$$y = a_4x' + a_5y' + a_6 \quad (3)$$

6개의 매개변수를 구하기 위해서는 3개의 점에 대응되는 좌표를 알고 있어야 한다. 이를 위해 삼각형 구조로 메쉬를 만들고 K frame에서 구한 각각의 3점으로부터 그림4 와 같이 K+1 frame 으로 Forward 예측하여 대응되는 점을 찾는다. 그후 6개의 매개변수를 구하고 Backward Affine 보상(Compensation)을 한다.

$$\begin{bmatrix} x_1' & y_1' & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1' & y_1' & 1 \\ x_2' & y_2' & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_2' & y_2' & 1 \\ x_3' & y_3' & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_3' & y_3' & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \end{bmatrix} \quad (4)$$

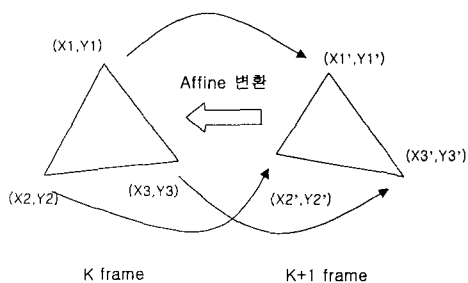


그림 4 Affine 변환

K+1 삼각형 내부의 점은 Affine 매개 변수만 알고 있으면 K frame 의 점으로 대응된다.

움직임 벡터(Displacement vector) 는 각 삼각 메쉬의 세점의 움직임 벡터로 다음과 같이 표현 할 수 있다.

$$\sum_{k=1}^3 d_k(p) = \begin{bmatrix} x_k' - x_k \\ y_k' - y_k \end{bmatrix} = \begin{bmatrix} u_k(x, y) \\ v_k(x, y) \end{bmatrix} \quad (5)$$

이때 (5) 와 같이 모션 벡터를 표현 할 수 있다면 한 메쉬의 Affine 변환한 식을 다음과 같이 정의 할 수 있다.

$$\sum_{p \in \mathbb{F}} A_f \left\{ \sum_{k=1}^3 d_k(p) \right\} \quad (6)$$

여기서 \mathbb{F} 는 삼각형 내부의 모든 이미지를 뜻하고 A_f 는 Affine 변환을 말한다.

노드 위치 변화에 따른 이미지 복구(Reconstruction)에 미치는 결과를 측정하기 위한 방법으로 모션 예측 에러(Motion prediction error)를 다음과 같이 정의 한다.

$$E_D = \sum_{p \in \mathbb{D}} (I(p + d(p), t + \delta t) - I(p, t))^2 \quad (7)$$

$$= \sum_{p \in \mathbb{D}} \sum_{p \in \mathbb{F}} (I(A_f(\sum_{k=1}^3 d_k(p) + p), t + \delta t) - I(p, t))^2$$

여기서 \mathbb{D} 는 Image 영역을 말한다

본 논문에서는 내용기반 메쉬를 생성하고 각각의 노드(Vertex)에서 모션 예측을 수행한다. 모션의 검색영역(Search area) 은 각각의 삼각 메쉬의 연결성(Connectivity)를 유지하기 위해 예측하는 현 노드가 다른 삼각 메쉬를 침범하지 않게 제한을 두었으며 ± 15 의 전 영역 탐색(Full search)를 행해서 MAD(Minimum mean absolute difference) 에러가 가장 작은 위치로 찾은 후 Affine 변환을 통한 역보상으로 영상의 재구성을 하였다.

IV. 실험 및 결과

실험에 사용된 영상은 360x288 크기의 이미지로 1~50 frame 중에서 이미지의 변화량이 많은 Frame을 중심으로 테스트를 수행한다. 실험 이미지는 Miss America 를 사용했고 알고리즘의 비교를 위한 기준으로 블록매칭을 3 step search 알고리즘을 사용해서 모션 예측을 한 후 k+1 frame을 복구하고 PSNR(Peak signal to noise ratio)를 구해서 메쉬 모델 예측 복원한 영상의 PSNR 과 비교하였다.

그림5 는 원 그림에서 에지 추출과 노드 선택과정을 거쳐 삼각 메쉬를 생성하는 과정을 보여준다.

그림6 은 40번째 frame을 각각 블록 매칭 방법과 메쉬 매칭 과정을 통해 복원한 그림을 보여준다.



그림 5 Triangulation 과정

그림 7 에서 볼 수 있듯이 블록 매칭의 경우는 움직임
 량이 많은 입 주위에 블록 노이즈가 발생하지만 메쉬
 매칭의 경우 경계선 사이에서 자연스러운 영상으로 복
 원됨을 알 수 있다.



그림 6 40 frame(좌) 41 frame(우)



그림 7 블록 매칭 복원(좌) 메쉬 매칭 복원(우)

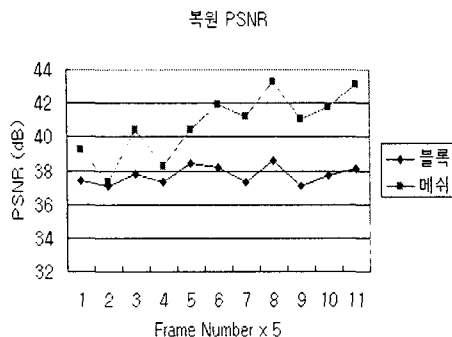


그림 8 각 방법 당 Frame 별 에러율 (PSNR)

그림8에서 볼 수 있듯이 블록 매칭 방법보다는 메쉬
 매칭 방법이 전체적 2 dB 정도 더 좋은 PSNR을 보여
 준다. 특히 모션이 많은 30 ~ 45 frame 에서는 더욱
 좋은 PSNR을 보여 준다.

결론적으로 블록 매칭에 비해 더 좋은 PSNR과 경계
 부분이 매끄러워지는 비주얼 효과를 볼 수 있었다.
 노드의 위치를 더욱 정밀하게 설정 한다면 더 좋은
 PSNR 과 비주얼 효과를 얻을 수 있을 것이다.

참고문헌

- [1] Kang W. Chun and Byeungwoo Jeon, "Irregular Triangular Mesh Representation Based on Adaptive Control Point Removal", SPIE volume 2727. pages 844-853, 1996.
- [2] M. Dudon ,O. Avaro, C. Roux. "Triangular active mesh for motion estimation", Signal proc.: Image Commun. vol 10, pages 21-41, 1997.
- [3] Yucel Altubasak and A. Murat Tekalp, "Occlusion-Adaptive, Content-Based Mesh Design and Forward Tracking", IEEE trans. Image processing, vol. 6, pages 1270-1280, September 1997.
- [4] Demin Wang, "Improvement of region-based motion estimation by considering uncovered regions", Signl proc. : Image commun. vol 14, pages 841-849, 1999.
- [5] Yao Wang and Jorn Ostermann, "Evaluation of Mesh-Based Motion Estimation in H.263-Like Coders", IEEE trans. Circuit and System for Video Technology, vol. 8, pages 243-252, June 1998.
- [6] A. Neri , S. Colonnese, "On the computation of warping-based motion compensation in video sequence coding", Signal proc. : Image commun. vol 13, pages 155-160, 1998.
- [7] Peter van Beek, Murat Tekalp, Ning Zhuang, "Hierarchical 2-D Mesh Representation, Tracking, and Compression for Object-Based Video", IEEE trans. Circuit and System for Video Technology, vol. 9, pages 353-369, March 1999.
- [8] ITU-T Recommendation H.261, Video codec for audio-visual services at p x 64 kbit/s, 1993.
- [9] S.W Sloan, "A fast algorithm for constructing Delaunay triangulations in the plane", Adv. Eng. Software, Vol. 9,No 1, pages 34-55, 1987.