

MPEG-4에 사용되는 동영상 객체면의 구성을 위한 레이블링과 블록 정합 방법

최 정 화(崔 正 和), 한 수 영(韓 秀 榮), 임 제 탁(林 濟 鐸)
한양대학교 전자공학과
전화 : (02) 2290-0353 / 팩스 : (02) 2298-1796

Labeling-BMA Algorithms for VOP of MPEG-4

Jung-Hwa Choi, Su-Young Han, Chae-Tak Lim
Dept. of Electronic engineering, Hanyang University
Tel. 02-2290-0353, Fax. 02-2298-1796
E-mail : jhchoi@hanlim.hanyang.ac.kr

Abstract

In this paper, we propose new algorithms to construct video object planes(VOP's) for MPEG-4. VOP's allow the new video standard MPEG-4 to enable content-based functionalities. A comprehensive review summarizes some of the most important VOP's generation techniques that have been proposed. The proposed algorithm use segmentation technique as labeling and motion estimation as three-step search algorithm(TSS). It is improved by a labeling technique that distinguishes background and object from a frame.

I. 서론

MPEG-1, MPEG-2, H.261, 그리고 H.263과 같은 동영상 기술은 장면의 분할(segmentation)과 같은 기술이 없는 저수준(low-level) 기술이었다. 이러한 기술들은 높은 압축율로 많은 응용분야에서 사용되고 있다. 하지만 점차 멀티미디어 분야의 인기가 증가함에 따라 사물에 기반한 새로운 동영상 기술이 필요하게 되었다.

새로운 동영상 부호화 표준인 MPEG-4의 특징은 객체(object) 기반 능력이 있다는 것이다. 이것은 입력된 동영상의 각 프레임들 안에 포함되어 있는 의미 있는 객체나 흥미 있는 객체와 그 외의 영역을 임의의 특징들로 분할된 다수의 동영상 객체평면(video object planes :VOP)을 구성함으로써 가능하다[1][2]. 즉, 동영상 객체층은 각각의 모양(shape), 움직임(motion), 질감(texture)등의 정보를 포함하는 동영상 객체 평면으로 구성된다.

의미 있다는 것을 무엇으로 볼 것인가[3][4]에 따라서 동영상 객체 평면을 나누는 방법이 다르게 될 것이다.

지금까지는 모양을 기반으로 하여 각 프레임을 분할하는 방법[5], 움직임을 기반으로 하여 분할하는 방법[6][7], 질감을 기반으로 하는 방법 중 하나의 특징만을 기반으로 하는 방법들이 제안되었다. 모양에 기반한 제안은 계산이 복잡하게 된다는 단점이 있다. 또한 움직임에 의한 방법은 객체들의 폐색(occlusion)과 같은 문제에 있어서 움직임의 정확한 측정이 어렵고 측정오차가 있을 경우에는 악영향을 받는다는 단점이 있다.

본 논문에서는 모양과 움직임을 기반하는 기법을 병행으로 사용함으로써 위의 단점을 보완하려고 한다. II에서는 모양에 기반을 둔 기법과 움직임에 기반을 둔 방법들에 대하여 간단히 요약한다. III에서는 제안한

방법에 대하여 상세히 설명한다. 그리고 IV에서 모의 실험 결과를 보인 후에 V에서 결론을 맺는다.

II. 모양(Shape), 움직임(Motion)에 기반한 방법들

모양에 기반한 방법은 한 프레임 안에 있는 영상물의 입의 매개 변수들로 표현하는 것이다. 이 방법으로는 경계선 추출(edge-detection)을 사용하여 특정 객체를 분할하는 것과 영역 분할(region partition), 그리고 왜곡선(contour)에 기초한 분할등이 있다[4]. 이러한 방법들은 계산량이 많다는 단점이 있다.

움직임(motion)에 의한 방법들을 간단히 다음과 같이 표현 할 수 있다.

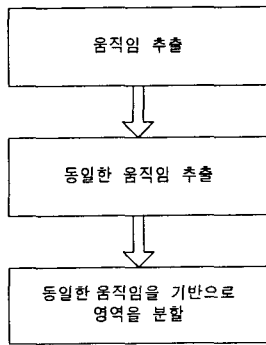


그림 1 움직임(motion)에 기반을 둔 VOP 구성

위 그림 1에서 보듯이 움직임에 기반한 방법들은 측정된 움직임에 의하여 분할하는 방법으로서 우선적으로 움직임을 정의해야 한다. 움직임을 표현하는 방법으로 움직임 정의방식에 따라 매개 변수 모형(Parametric model)과 비매개 변수 모형(Non parametric model)으로 나눌 수 있다. 매개 변수 모형(Parametric model)으로는 Affine 모형[8]과 Perspective 모형등이 있다. Affine 모형은 여섯 개의 매개 변수 $\{a, b, c, d, e, f\}$ 로써 움직임을 다음과 같이 나타낼 수 있다.

$$\begin{aligned} x_{n-1} &= T_x(x_n, y_n) = ax_n + by_n + c \\ y_{n-1} &= T_y(x_n, y_n) = dx_n + ey_n + f \end{aligned} \quad (1)$$

AFFINE 모형의 방정식 식(1)을 계산하기 위해서는 최소한 3점이 필요하게 된다. AFFINE 모형은 전역 또는 지역적인 움직임을 표현하는데 사용한다.

Perspective 모형은 여덟 개의 매개 변수 $\{a, b, c, d, e, f, g, h\}$ 가 필요하며

$$\begin{aligned} x_{n-1} &= T_x(x_n, y_n) = \frac{ax_n + by_n + c}{gx_n + hy_n + 1} \\ y_{n-1} &= T_y(x_n, y_n) = \frac{dx_n + ey_n + f}{gx_n + hy_n + 1} \end{aligned} \quad (2)$$

나타낼 수 있다. Perspective 모형의 방정식 (2)를 위해서는 최소한 4점이 필요하게 된다. AFFINE 모형과 perspective 모형등의 매개 변수 모형은 반복적으로 매개 변수를 변화 시킴으로써 방정식에 만족하는 값을 구해야 한다. 따라서 많은 계산량이 필요하게 된다. 비매개 변수 모형으로는 블록 정합(Block matching) 방법이 있다. 움직임에 기반한 방법들은 측정된 움직임(motion)을 기반으로 유사한 움직임을 갖는 영역 단위로 분할(segmentation)한다.

III. Labeling-Motion Estimation

제안한 방법에 대하여 자세히 설명도록 한다. 의미영역을 I-frame을 먼저 분할하고 그 분할된 영역을 기반으로 움직임을 추정하는 방법이다.

1. labeling 된 영상 구현

먼저 $F(k-1)$ 번째 프레임의 전체 픽셀값(resolution) $S(x, y, k-1)$ 를 분석한다.

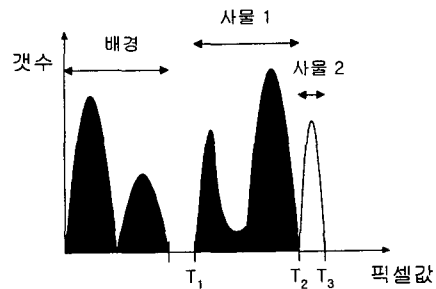


그림 2 $F(k-1)$ 번째 프레임의 $S(x, y, k-1)$ 의 분포도

$$T_n \leq S_{Cn}(x, y, k-1) \leq T_{n+1}$$

$$T_n > S_{NCn}(x, y, k-1) \text{ or } S_{NCn}(x, y, k-1) > T_{n+1}$$

여기에서 $S_{Cn}(x, y, k-1)$ 는 관심 있는 객체 영역의 픽셀값을 의미하고 $S_{NCn}(x, y, k-1)$ 은 그 외의 영역들 특히 배경등에 해당하는 픽셀값을 의미한다. 표현

가능한 최고 해상도 $S_{\max}(x, y, k-1)$ 와 관심 있는 객체의 수 n 을 정한다. 그리고 단위 레이블값(unit label value) l_{nk-1} 를 정한다.

$$l_{nk-1} = \frac{S_{\max}}{n+1} \quad n = 1, 2, 3, \dots$$

객체에 대한 관심도에 따라서 객체에 가중치가 부여된 객체수 n 을 설정한다.

$$L_{nk-1}(O_n) = l_{nk-1} \times (n+1)$$

여기에서 O_n 은 가중치가 부여된 객체수 n 을 의미하고 $L_{nk-1}(O_n)$ 은 가중치 n 객체에 해당하는 레이블(label)값을 표현한다. 관심이 없는 객체와 배경(background)의 $L_{0k-1}(O_0)$ 는 0이다.

$F(k)$ 번째 프레임의 전체 픽셀값(resolution) $S(x, y, k)$ 를 분석하고 $S(x, y, k-1)$ 와 유사하면 $F(k-1)$ 의 레이블값들을 사용하여 레이블된 $F'(k)$ 를 표시한다.

$F(k+1)$ 번째 프레임도 마찬가지로 $S(x, y, k+1)$ 도 해석한다. $F(k+1)$ 번째 프레임도 유사한 형태 즉, 히스토그램의 형태가 유사하다면 $F(k)$ 프레임의 단위 레이블값(unit label value)을 기준으로 하여 레이블된 $F'(k+1)$ 을 구성한다.

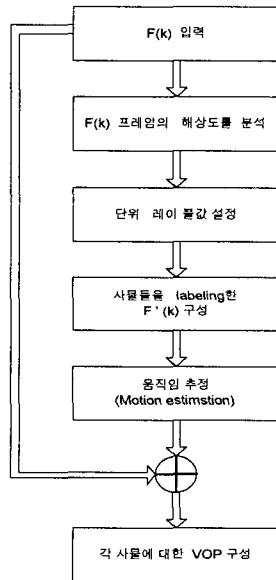


그림 4. 제안된 기법의 순서

2. 움직임 측정

움직임 측정은 블록 정합(BMA)에 근거한다. 원영상 $F(k)$ 와 $F(k+1)$ 이 아닌 레이블화(labeled)된 영상 $F'(k)$ 과 $F'(k+1)$ 을 탐색 대상으로 정한다.

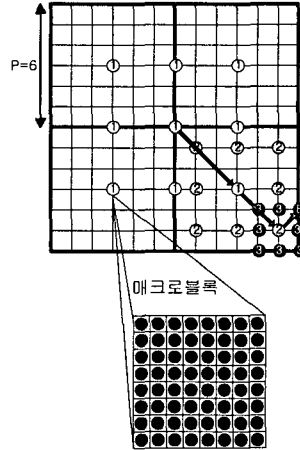


그림 3. 3-단계 블록 정합 기법(3-step Search)

블록 정합(BMA) 방법 중에서 탐색 시간을 줄이기 위하여 전역 탐색(full search)이 아닌 빠른 기법(fast algorithm)인 3단계 블록 병합(3-Step Block search : TSS)를 사용한다[9].

그림 3과 같이 탐색 영역 계수(search area parameter)는 6으로 설정하고 매크로블럭(macroblock)의 크기는 8×8 이다.

IV. 모의 실험

실험에 사용된 영상은 352×240 크기로 10개의 프레임을 사용하였다.



그림 5. 사용된 sample 영상 $F(k)$ 와 $F(k+1)$

$F(k-1)$ 의 히스토그램을 기반으로 하여 관심 있는 객체 수와 레이블 값을 설정한다.

실험에서는 탁구공에 관심을 두었으며 객체수 n 은

1 이므로 탁구공에 레이블값(255)이 할당 되었다.

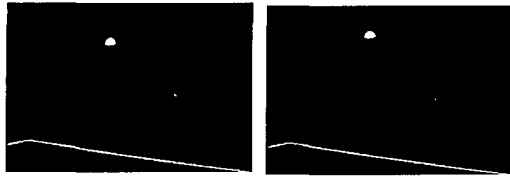


그림 6. 레이블된 영상 $F'(k)$ 와 $F'(k+1)$

그림 6의 레이블화된 영상 $F'(k)$ 와 $F'(k+1)$ 을 기반으로 하여 움직임을 추정할 것이다. 움직임 추정에 사용된 기법으로 3-단계 탐색(3-Step search)을 사용하였다.

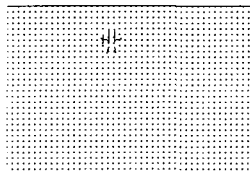


그림 7. $F'(k)$ 와 $F'(k+1)$ 사이의 움직임 흐름도

그림 7은 $F'(k)$ 와 $F'(k+1)$ 사이의 움직임을 측정 한 것이다. 10개의 프레임을 연속적으로 객체(탁구공)만의 움직임을 추출하고 복원하였을 때의 PSNR이다.

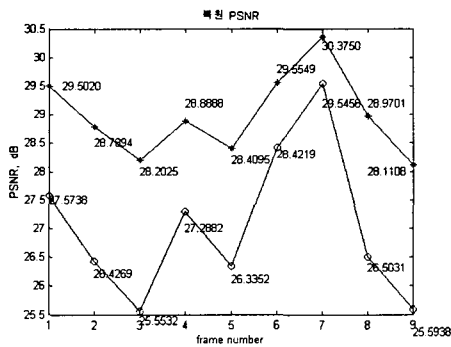


그림 8. TSS(*)와 제안된 기법(o)의 복원 PSNR

실험 결과에서 움직임 추정은 관심 객체에만 한정 되었으며 객체의 움직임만으로 복원된 영상의 PSNR은 전체 움직임을 복원한 PSNR과 거의 유사하였다.

V. 결론

본 논문에서 제안된 방법은 동영상의 새로운 표준인

MPEG-4의 객체지향성을 추구하기 위하여 동영상 객체층을 구성하기 위한 방법이다. 객체의 선택이 픽셀 값에 있다. 따라서 시각에 민감한 방법으로 객체를 선택하는 방법이 간단하고 움직임 측정 시간이 빠르다는 장점이 있다. 관심 있는 객체들이 그 외의 영역과 픽셀값에서 차이가 클수록 더욱 효과적이다.

참고문헌

- [1] Thomas Meier and King N. Ngan "Automatic Segmentation of Moving Objects for Video Object Plane Generation" IEEE Trans. on Circ. and Syst for Video Tech. VOL.8, No. 5, pp. 525-537, Sep 1998.
- [2] Yining Deng and B. S. Manjunath "NeTra-V: Toward an Object-Based Video Representation" IEEE Trans. on Circ. and Syst for Video Tech. VOL.8, No. 5, pp. 616-627, Sep 1998.
- [3] A. M. Tekalop, Digital Video Processing. Prentice-Hall, 1995
- [4] Milan. Sonka, Image Processing Analysis and Machine Vision, ITP, 1998.
- [5] W. Y. Ma and B. S. Manjunath "Edge flow: A framework of boundary detection and Image segmentation," in Proc. IEEE Conf. Comput. Vision Pattern Recognition, pp. 744-749, 1997.
- [6] Georgi D. Borshukov. Gozde Bozdagi, Yucel Altunbasak, and A. Murat Tekalp "Motion Segmentation by Multistage Affine Classification" IEEE Trans. on Image Processing, VOL. 6, NO. 11, pp. 1591-1597, Nov 1997.
- [7] Etienne Mémin and Patrick Pérez "Dense Estimation and Object-Based Segmentation of the Optical Flow with Robust Techniques" IEEE Trans. on Image Processing, VOL. 7, NO. 5, pp. 703-719, May 1998.
- [8] Zhaohui Sun and A. Murat Tekalp "Trifocal Motion Modeling for Object-Based Video Compression and Manipulation" IEEE Trans on Circ and Syst for Video Tech. VOL. 8, No. 5, pp. 667-685, Sep 1998
- [9] Jianhua Lu and Ming L. Liou "A Simple and Efficient Search Algorithm for Block-Matching Motion Estimation" IEEE Tran. on Circ and Syst for Video Tech. VOL. 7, No. 2, pp. 429-433, Apr 1997.