대화형 방송을 위한 객체위치 추적 시스템

안준한, 고재필, 변혜란
연세대학교 컴퓨터과학과

Object-Tracking System For Interactive Broadcasting

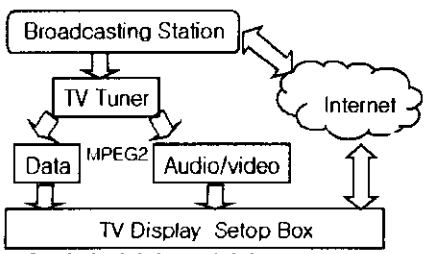
Junhan Ahn, Jaepil Ko, hyeran Byun
Dept. of Computer Science, Yonsei University
E-mail : foryou@yonsei.ac.kr

요약

디지털 TV 가 실용화 됨에 따라 다양한 부가정보 서비스가 가능하게 되었다. 그러나 부가정보서비스를 효과적으로 사용하기 위해서는 전통적인 메뉴검색에 의한 관심객체의 부가정보 검색이 아닌 화면에서의 관심객체 선택만으로 부가정보를 표현할 수 있는 방법이 필요하다. 따라서 본 논문에서는 메뉴검색 없이 마우스 클릭만으로 관심 객체를 선택하고 선택된 객체에 대해 부가정보를 표현하는 시스템을 제안한다.

1. 서론

2001 년 9 월 디지털지상파 방송을 앞두고 방송용 컨텐츠 산업이 활기를 띄고 있다. 기존의 아나로그 TV 와 PC 를 접목한 WebTV 가 96 년 본격 시판된 이후 5 년이 지난 지금 다양한 부가정보를 포함할 수 있는 디지털 TV 와 PC 의 접목이 새로이 시도 되고 있다. 기존의 WebTV 는 아나로그 TV 와 PC 가 하나의 시스템으로 결합되었지만 실질적으로 TV 와 PC 가 독립적으로 작동하는 시스템 이였다. 그러나 대화형 TV 는 디지털지상파 방송에 부가정보를 포함시키고 이것에 대한 시청자의 반응을 셋톱박스를 이용하여 즉각적으로 반응 할 수 있는 시스템이다. 이러한 대화형 TV 의 등장으로 TV 프로그램의 채널별, 시간대별 안내는 물론, 드라마에서의 등장인물에 대한정보, 의상, 배경장소, 다큐멘터리에서의 상세정보와 용어해설 등의 부가정보를 제공 할 수 있을 뿐만 아니라 시청자가 관심 있는 상품을 TV 시청중 바로 구매할 수 있는 E-Commerce(Electric Commerce)도 가능하다. 드라마 시청 중 주인공의 옷이나 소품에 관심이 있는 경우 몇 번의 메뉴검색 만으로 해당물품을 구매할 수 있도록 하는 것이 그 예이다.



[그림 1] 대화형 TV 데이터 흐름도

2. 연구 범위

본 논문에서는 대화형 방송환경에서의 부가정보 삽입과 부가정보 표현 방법 중 사용자가 관심 있는 객체를 직접 선택했을 때 선택한 객체에 대한 부가 정보를 표현하고 즉시 관련 사이트로 연결하는 시스템을 구현하는 것이다. 이를 위해서는 디지털방송에서 사용되는 MPEG 스트리밍 파일을 편집 가능한 프레임 단위로 나누고 선택한 객체를 트랙킹 하는 기술이 필수적이다 따라서 본 논문에서는 Microsoft 에서 제공되는 DirecX 8.0 을 기반으로 스트리밍 파일을 독립적인 프레임으로 나누고 각각의 프레임에서 선택한 객체를 추적하는 시스템을 구현하였다.

3. 객체추적 방법

비디오 데이터에서 부가정보 삽입의 대상이 되는 특정 객체를 추적하기 위한 방법은 크게 움직임기반 추적과 모델기반 추적 두 가지 접근 방법을 사용하고 있는데, 실제 비디오에서 부가정보 삽입을 위한 객체는 그 종류가 너무 많아서 모든 객체에 대해 모델을 만들 수 없기 때문에 본 시스템에서는 대상 객체를 크게 사람과 사물로 나누어, 사람의 얼굴은 모델을 만들어 추적하는 모델기반 추적방법을 사용하고 나머지 사물에 대해서는 객체의 영역을 지정하여 영역을 추적하는 움직임기반 추적을 적용하였다. 우선 움직임기반 추적은 이전프레임과 현재프레임의 차이를 계산하여 움직임을 검출 하고 검색영역 중에서 객체를 찾는 방법이다. 모델기반 얼굴 추적 방법에서는 각 프레임에서 독립적으로 얼굴을 찾아내는 방법으로 이때 사용하는 방법으로는 크게 색상 모델과 모양모델(타원)을 이용한다. 그러나 색상 모델만을 이용한 방법은 조명,배경에 따라서 많은 오류를 나타내고 얼굴이 회전하여 뒷모습이 나타날 때는 계

속적으로 추적할 수 없다는 단점을 가지고 있고 모양모델만을 이용한 방법은 영상에서 타원과 유사한 모양의 모델이 나타나면 잘못된 추적이 될 수 있기 때문에 본 시스템에서는 색상 모델과 모양모델을 동시에 적용하였다.

3.1 움직임 기반 추적방법

특정 객체의 영역을 추적 하기 위한 방법으로는 블록정합방법[1] 중 검색 지점의 개수를 대수적으로 줄이는 Three-step search algorithm[2]을 적용하였다. Three-step search algorithm 은 첫번째 단계에서 자신의 위치와 검색영역크기의 절반 지점인 8 개 지점을 선택하여 비교하고 가장 유산한 한 개의 지점을 선택한다. 두번째 단계에서는 첫번째 단계에서 선택된 지점을 다시 검색영역크기의 절반 지점인 8 개 지점을 선택하여 비교한다. 예를 들어 검색 영역의 크기가 ±7 화소 이면 비교하여야 할 지점은 25(9+8+8)가 된다. 그러나 일반적인 블록정합 방법에서는 검색영역크기가 ±7 화소 이기 때문에 장면의 진행이 급격하게 변하는 비디오 데이터에서는 객체의 영역을 놓치는 경우가 발생한다. 이러한 문제점을 극복하기 위하여 검색영역크기를 확대할 수도 있지만 검색영역크기 확장은 곧 계산량의 증가로 이어져서 전체 시스템의 수행속도에 많은 영향을 미치게 된다. 따라서 본 시스템에서는 공간적 예측을 이용한 Three-step search 방법에 시간적 예측(Temporal Prediction)을 함께 적용하여 검색영역크기 확장 없이 ±7 화소 이상의 변화까지 해결하였다(그림 1 참조). 시간적 예측방법은 이전 프레임들에서의 객체위치 변화량 데이터로부터 다음 프레임에서의 객체위치를 예측하고 다음 프레임의 검색영역을 예측된 위치로 이동하여 검색하는 방법이다. 즉 객체영역의 중심좌표가 2,7,13,20,28 의 순서로 이동하고 있다면 다음 중심위치는 37 이 될 것이고 검색영역의 위치는 28 에서 ±7 이 아닌 37 에서 ±7 이 된다. 다음 수식(1)은 이러한 중심좌표의 예상을 간단히 모델링 한 것이다.

$$x_t^p = 3x_{t-1} - 3x_{t-2} + x_{t-3}$$
$$y_t^p = 3y_{t-1} - 3y_{t-2} + y_{t-3} \quad (1)$$

간단한 수식(1)의 예측만으로도 정지, 가속운동, 등속운동, 감속운동을 효과적으로 예상하여 검색영역을 정할 수 있다.

그림 1 객체의 움직임이 ±7 화소 이상 움직인 경우

그림 1 에서 추적하기를 원하는 객체영역은 두개의 의자이며 검색영역은 객체영역을 포함하고 있는 외각의 큰 사각형이다. 그림 1 비디오 영상에서의 객체 움직임은 9 픽셀로써 검색영역을 초과하기 때문에 단순한 Three-step search 방법으로는 실패하였다. 그러나 시간적 예측을 적용한 Three-step search 방법에서는 이전 프레임에서 객체 움직임을 측정하고 측정된 데이터로부터 다음프레임에서의 객체위치를 예상하기 때문에 검색영역의 크기 증가 없이 효과적인 추적이 이루어 졌다. 다음 표 1 은 범용적으로 사용되는 테니스영상과 가든영상을 대상으로 3SS(Three-step Search)와 TP3SS(Temporal Prediction Three-step Search)의 평균 MAE(Mean Absolute Error)의 차이를 비교한 것이다

표 1 처음프레임부터 50 프레임까지의 MAE

Sequence	3SS	TP3SS
Tennis	6.876875	6.870781
Garden	13.372265	13.370976

검색영역크기 6

Sequence	3SS	TP3SS
Tennis	7.481914	7.110156
Garden	15.075507	13.895546

검색영역크기 4

표 1 에서 테니스영상과 가든영상은 영상의 움직임이 대부분 6 화소 이내에서만 움직이기 때문에 검색영역크기가 6 화소일 경우 시간적 예측방법을 사용하여도 MAE 의 변화가 거의 없다. 하지만 검색영역크기가 4 로 줄어들면 시간적 예측방법을 사용한 추적방법이 더 정확하게 추적되었음을 알 수 있다.

3.2 모델기반 얼굴 추적방법

각각의 프레임 영상에서 얼굴사이즈와 위치는 얼굴모델에 의해서 결정된다. 얼굴모델은 단축과 장축의 비가 1:1.3 인 타원모델[3]과 타원의 내부 색깔이 피부색에 해당하는지를 조사하는 색상모델[4]을 결합하여 생성한다(그림 2 참조).

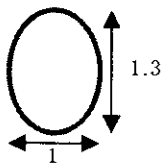


그림 2 얼굴모델을 위한 장축과 단축의비

3.2.1 타원모델

타원모델은 단축의 길이와 장축의 비가 1:1.3 인 타원을 만들고 원 영상의 이진화 영상을 생성하여

이진화 영상에서 타원모델과 가장잘 정합 되는 타원모델의 위치와 크기를 찾아내는 것이다. 즉 단축의 크기가 40 화소인 타원부터 20 화소인 타원모델을 이진화 영상의 검색영역에서 모두 정합 시켜보고 가장 잘 일치하는 타원을 얼굴영역으로 판단 하는 것이다. 여기서 타원모델을 형성하고 있는 픽셀의 위치에 이진화 영상의 픽셀값이 1 이면 Score 의 값이 1 증가한다. 타원모델이 가장 잘 정합 된다는 것은 타원모델의 전체 픽셀개수 중에서 몇 개의 픽셀이 이진화 영상과 교차 시켰을 때 1 을 나타내는지를 살펴보고 그 개수가 가장 큰 것이 가장 잘 정합 되는 위치와 크기로 결정되는 것이다.

$$Score = \max_{s \in S}\left\{\frac{1}{n_\delta}\sum_{i=1}^{n_\delta} |\, Gredient_i \,|\right\} \quad (2)$$

$S=(x,y,\delta)$ x,y: 는 타원 중심위치 δ : 타원 단축 길이
최초의 얼굴 중심위치는 시스템 사용자에 의해서 초기화 되며, 검색 영역은 ±7 화소이다.

크기가 40 화소인 타원 (score = 20)

크기가 30 화소인 타원 (score = 160)

그림 3

3.2.2 색상모델

$$Score = \max_{s \in S}\left\{\frac{1}{I_\delta}\sum_{i=1}^{I_\delta} FR_i\right\} \quad (3)$$

FR(x,y)=1, if $H \le T_1 \bigcap S \le T_2$
=0, otherwise
FR:타원내부영역, H:휘도(Hue), S:채도(Saturation)

색상모델만으로도 제한된 환경에서는 얼굴을 추적 할 수 있다. 하지만 컬러모듈은 조명과 얼굴 색상에 민감한 반응을 보이고 얼굴이 회전하면 색상 정보가 사라지기 때문에 추적에 실패한다. 따라서 이러한 단점을 극복하기 위해 본 시스템에서는 타원모델과 색상모델을 동시에 적용하였다(수식 4).

$$Score = \max\{Ellipse\_score + \lambda Color\_score\} \quad (4)$$

λ는 타원모듈에 대한 컬러모듈의 가중치이다

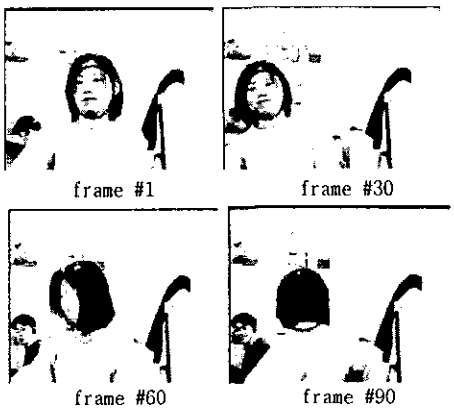


그림 4 타원모델과 컬러모델을 결합한 결과

5.결론 및 향후 연구방향

본 논문에서는 대화형방송환경에서 메뉴검색방식으로 부가정보를 검색하는 방법이 아니라 영상에서의 객체 클릭에 의한 객체 부가정보를 표현하는 방법에 대하여 제안하였다. 그러나 지상파 방송에서 사용되는 영상들은 카메라의 줌인(Zoom in)이나 줌아웃(Zoom out)에 의하여 객체의 크기가 지속적으로 변화되고 있다. 이러한 문제점을 해결하기 위해서는 움직임 기반 추적방법에서 고정된 크기의 움직임 추적이 아닌 가변적 크기의 움직임을 추적 하는 방법이 필요하며, 이를 위한 연구가 진행중이다.

6. 참고문헌

[1] J. B. Xu, L. M. Po and C. K. Cheung, "Adaptive motion tracking block matching algorithms for video coding," *IEEE Trans. on Circuits Syst. Video Technol.* vol. 97, Oct.,1999, pp.1025-1029

[2] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, " Motioncompensated interframe coding for video conferencing," Proc. NTC81, pp. G5.3.1, Nov. 1981.

[3] S. Birchfield. An elliptical head tracker. In *Proc. of the 31st Asilomar Conf. on Signals, Systems and Computers*, 1997.

[4] C. Bregler. Learning and recognizing human dynamics in video sequences. In Proceedings of IEEE CVPR 97, page 568-574, 1997.

[5] Jianbo shi, Carlo Tomasi, Good Features to Track. In IEEE conference on computer vision and Pattern Recognition (CVPR94) Seattle, June 1994.