대화형 방송 환경에서 부가서비스 제공을 위한 객체 추적 시스템

(Object Tracking System for Additional Service Providing under Interactive Broadcasting Environment)

안 준 한 † 변 혜 란 ††
(JunHan Ahn) (Hyeran Byun)

요 약 본 논문은 대화형 방송환경에서 부가서비스를 제공받기 위해서 탑다운(Top-Down)메뉴 검색을 하는 것이 아니라, 방송영상의 화면 내부에서 부가서비스가 제공되길 원하는 객체를 선택했을 때 선택한 객체에 대한 부가서비스를 제공하는 새로운 방법을 제안한다. 이를 위해서는 실시간으로 방송되고 있는 동영상과 객체정보(위치, 크기, 모양)의 동기를 맞추는 기술과 동영상 내부의 객체 추적 기술이 필수적이다. 동영상과 객체정보의 동기를 맞추는 기술은 마이크로소프트사의 다이렉트쇼(DirectShow)를 이용하였으며, 객체를 추적하기 위한 방법은 객체를 크게 사람과 사물로 나누어, 사람의 얼굴은 모델을 만들어 추적하는 모델 기반 얼굴 추적 방법(Model-based face tracking)을 사용하고 나머지 사물에 대해서는 객체의 영역을 지정하여 영역을 추적하는 움직임 기반 추적 방법(Motion-based Tracking)을 적용하였다. 또한 움직임 기반 추적 방법에는 시간적 예측 검색 방법을 적용하여 움직임이 큰 객체도 검색 영역 확장 없이 정확한 추적을 할 수 있도록 하고 모델 기반 추적 방법에는 타원 모델과 색상 모델을 결합한 얼굴 모델을 적용하여 얼굴이 회전하여도 정확한 추적을 할 수 있도록 개선하였다.

키워드 : 데이타 방송, 대화형 방송, 객체 추적

Abstract In general, under interactive broadcasting environment, user finds additional service using top-down menu. However, user can't know that additional service provides information until retrieval has finished and top-down menu requires multi-level retrieval.

This paper proposes the new method for additional service providing not using top-down menu but using object selection. For the purpose of this method, the movie of a MPEG should be synchronized with the object information(position, size, shape) and object tracking technique is required. Synchronization technique uses the Directshow provided by the Microsoft. Object tracking techniques use a motion-based tracking and a model-based tracking together. We divide object into two parts. One is face and the other is substance. Face tracking uses model-based tracking and Substance uses motion-based tracking base on the block matching algorithm. To improve precise tracking, motion-based tracking apply the temporal prediction search algorithm and model-based tracking apply the face model which merge ellipse model and color model.

Key words : Interactive broadcasting, Data broadcasting, Object tracking

1. 서 론

멀티미디어의 기술 발달과 인터넷의 대중화로 최근 문자뿐만 아니라 오디오 및 비디오를 포함하는 멀티미디어 데이타에 대한 수요와 공급이 급증하게 되었고 이런 멀티미디어 데이타의 효율적인 저장과 전송을 위해 활발한 연구가 진행되고 있다. 특히 디지털 방송과 영상통신(visual communication)에 관계된 MPEG-2 (Motion Picture Expert Group)의 급속한 발전으로 현재 이러한 기반기술을 이용한 디지털 지상파 방송이 현실로 다가왔다.

아날로그 방송에서는 6MHz 대역폭에 영상과 음향신

† 학생회원 : 연세대학교 컴퓨터과학과
foryou@aipiri.yonsei.ac.kr
†† 종신회원 : 연세대학교 컴퓨터과학과 교수
foryou@yonsei.ac.kr
논문접수 : 2001년 7월 2일
심사완료 : 2001년 10월 24일

호로 구성된 하나의 TV 프로그램밖에 보내지 못했지만, MPEG-2 기반기술을 이용하면 같은 6MHz 대역폭에 서너 개 이상의 TV 프로그램을 압축하여 보낼 수 있으며 여기에 부가서비스를 위한 데이타도 추가하여 보낼 수 있다. 이러한 부가서비스의 삽입으로 기존에는 불가능했던 다양한 서비스가 이루어지게 되었다. 예를 들면 방송중인 프로그램에 관련된 정보 또는 주식, 환율, 일기예보 등 다수의 실시간 정보, 인터넷 정보 등 멀티미디어 부가서비스가 가능해진 것이다. 뿐만 아니라 지금의 아날로그 방송처럼 단방향 통신이 아니라 양방향 통신이 가능한 대화형 TV가 등장하게 되었다.

대화형 방송은 방송사로부터 기본적인 뉴스, 날씨, 프로그램 관련정보 등 각종 대화형 부가서비스를 제공받고, 전화선이나 별도의 전용선을 통해 부가서비스에 삽입된 연결정보로 부가서비스 제공업체에 접속하여 더 많은 정보를 제공받게 되는 것이다.

대화형 방송의 부가서비스는 TV프로그램의 채널별, 시간대별 안내(Electric Program Guide)뿐만 아니라 드라마나 쇼에서 등장인물에 대한정보, 의상, 촬영장소, 다큐멘터리에서의 상세 정보와 용어해설 등의 부가정보를 제공 할 수 있다. 또한 시청자가 관심 있는 상품을 TV 프로그램 시청 중 바로 구매할 수 있는 T-Commerce (Television Commerce)도 가능하다. 예를 들어 드라마 속의 주인공이 입고 있는 옷에 관하여 관심이 있을 경우 영상과 함께 제공되는 부가서비스의 검색을 통하여 옷에 관한 상세 정보를 얻을 수 있고, 부가서비스를 제공한 업체의 웹사이트로 바로 연결되어 구매도 할 수도 있다.

대화형 방송에서 시청자는 원하는 부가서비스를 찾기 위해 일정한 시간 간격을 두고 새롭게 갱신되는 부가서비스를 전통적으로 사용되는 탑다운(Top-Down) 메뉴 검색 방식을 통하여 검색한다. 하지만 이러한 탑다운 검색방식은 실제 시청자가 원하는 객체의 부가서비스를 포함하고 있지 않는 경우도 있으며 여러 단계의 버튼 검색을 해야하는 불편함이 있다. 이러한 문제점을 극복하기 위해서는 전통적인 메뉴 검색 방식이 아니라 방송되고 있는 영상의 화면 내부에서 원하는 객체를 선택했을 때 선택한 객체의 부가서비스를 제공하고 부가서비스를 제공한 업체의 웹사이트로 연결할 수 있어야 한다. 이를 위해서는 실시간으로 방송되고 있는 멀티미디어 스트림 데이타와 부가서비스를 위한 객체정보(위치, 크기, 모양)의 동기를 맞추는 기술과 동영상 내부의 객체 추적 기술이 필수적이다. 따라서 본 논문에서는 동영상을 각각의 프레임으로 나누고 각각의 프레임에서 선택한 객체의 위치를 추적하는 시스템을 구현하였다.

2. 관련연구와 연구범위

2.1 대화형 방송

대화형 방송 시스템은 디지털 지상파 방송에 부가서비스를 포함시키고 이것에 대한 시청자의 반응을 셋톱박스(set-top-box)를 이용하여 즉각적으로 반영 할 수 있는 시스템이다[1][2][3].

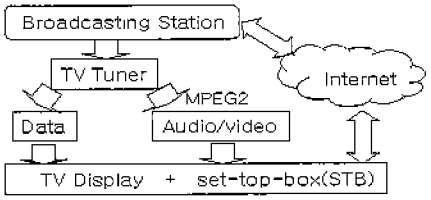


그림 1 대화형 방송의 데이타 흐름도

이러한 대화형 방송이 가능한 것은 방송사가 영상뿐만 아니라 부가서비스를 위한 정보를 함께 시청자에게 송출한다. 따라서 시청자는 TV시청 중 원하는 부가서비스를 검색하여 원하는 정보를 얻을 수 있으며 시청중인 프로그램에 대한 반응을 셋톱박스(set-top-box)를 이용하여 다시 방송국에 보낼 수 있다. 또한 부가서비스에서 제공되지 않은 더 많은 정보를 원하는 경우에는 부가서비스에 포함된 링크 정보를 이용하여 원하는 정보를 얻을 수 있는 것이다. 대화형 방송이 제공하는 서비스를 정리하면 다음과 같다[4].

- 프로그램 안내정보
  채널별, 시간대별, 주제별 프로그램 안내 및 프로그램 예약서비스
- 프로그램 관련 정보
  드라마(등장인물, 배경음악, 촬영장소), 다큐멘터리(상세정보, 용어해설, 장소설명)
  스포츠(경기전적, 일정, 선수소개)
- 대화형 서비스
  시청자 참여 퀴즈 프로그램, 실시간 여론조사
- TV 전자상거래(T-Commerce)

그림 2는 방송 시청 중 어떻게 시청자가 부가서비스를 이용하는지 보여주고 있다. 왼쪽그림에서 왼쪽 하단에 나타나는 "I"문자는 현재 진행중인 방송에 대화형 부가서비스가 삽입되어있다는 것을 나타내고 있으며, 오른쪽 하단의 대화형 버튼을 선택하면 오른쪽과 같은 화면으로 전환되어 시청자는 현재 방송되고 있는 장면에 대한 상세한 부가정보를 제공받을 수 있을 뿐만 아니라 온라인으로 상품 구매도 할 수 있다[5].

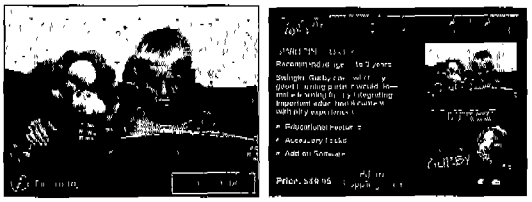


그림 2 일반적인 방송화면과 부가서비스 검색을 위한 화면

2.2 객체 추적

비디오 데이타에서 객체를 추적하는 방법은 크게 움직임 기반 방법과 모델 기반 방법으로 나뉘어지는데 움직임 기반 방법으로는 화소 차이값 분석 방법, 배경 보정 방법, 블록 정합 방법 등이 있고 모델 기반 방법에는 색상 모델과 모양 모델 방법 등이 있다.

- 화소 차이값 분석 방법(Temporal Differencing method)

카메라가 고정된 경우는 배경(Background)이 고정되어있으므로 간단히 다음 프레임에서 이전 프레임의 화소값을 빼면 고정된 배경은 모두 제거되고 움직임이 발생한 위치의 화소값만 남게 된다. 이 중 임계값 이상으로 화소값이 남아 있는 영역이 움직인 영역으로 판단되는 것이다[6].

$$\triangle_n = | I_n - I_{n-1} |$$

$$M_n(x,y) = \begin{cases} I_n(x,y), & \triangle_n(x,y) \geq T \\ 0, & \triangle_n(x,y) < T \end{cases} \quad (1)$$

I_n은 n^{th}프레임의히스토그램 $\triangle$은 difference function

M은 움직인 영역, 색상의 숫자가 255인 경우 T≈40

- 배경 보정 방법(Background Compensation)

카메라가 움직이는 경우는 배경이 움직인 것과 동일한 현상이 발생하므로 카메라 동작에 따른 배경 움직임을 보정해 주어야 한다. 카메라의 움직임을 보정해 줌으로써 배경은 고정된 것으로 계산되고 배경에서 실제로 움직임이 발생한 영역을 찾아내는 방법이다. 이러한 카메라의 움직임 보정 방법으로 많이 사용되는 방법은 카나타니스 포뮬러(Kanatani's formula) 이다[7].

$$x_{t-1} = f\frac{x_t + \alpha\sin\theta\, y_t + f\alpha\cos\theta}{-\alpha\cos\theta\, x_t + \gamma\, y_t + f}$$
$$y_{t-1} = f\frac{-\alpha\sin\theta\, x_t + y_t - f\gamma}{-\alpha\cos\theta\, x_t + \gamma\, y_t + f} \quad (2)$$

α pan angle, γ tilt angle, θ initial inclination of the camera

(x_t, y_t) 다음프레임에서의 화소의 위치

(x_{t-1}, y_{t-1}) 이전프레임에서 화소의 위치

- 블록 정합 알고리즘(Block Matching Algotithms)

블록 정합 알고리즘은 비디오 데이타내의 동작 정보를 추출하기 위해 사용되는 기본적이면서도 중요한 기술이다. 여기서 동작 정보는 구체적으로 카메라 동작에 따른 배경 영역의 움직임과 이동 물체의 움직임으로 구분할 수 있다. 객체의 움직임 정보는 블록 정합 알고리즘을 적용해 얻어진 동작 벡터를 분석함으로써 비디오 영상내의 동작 정보를 추출한다. 블록 정합 알고리즘의 종류에는 전역 탐색 알고리즘(FSBMA : Full Search BMA), 전역 탐색의 단점인 높은 시간적 복잡도를 줄이기 위해 만들어진 3단계 탐색 알고리즘(TSS : Three Step Search) 과 탐색 공간을 적절하게 조절해 나가는 적응적 전역 탐색 알고리즘(AFSBMA : Adaptive FSBMA) 등이 있는데 계산량은 많지만 가장 정확하게 검색되는 전역 탐색 알고리즘을 많이 사용한다.

- 색상 기반 추적 방법(Color-based Tracking)

색상 기반 추적 방법들은 추적하고자 원하는 객체의 색상 모델을 미리 만들어놓고 영상에서 색상 모델과 가장 잘 정합 될 수 있는 영역의 위치와 크기를 결정하는 방법이다. 컴퓨터에서 사용하는 RGB 색상 구조를 HSV 또는 YIQ 색상 모델로 변화하여 각각의 색상 모델과 가장 잘 정합되는 영역을 찾는 방법이다[8][9][10][11].

- 모양 기반 추적 방법(Shape-based Tracking)

모양 기반 추적 방법들은 추적하고자 원하는 객체의 2-D, 3-D모델을 생성한 다음 검색 영역에서 모양 모델과 가장 잘 정합될 수 있는 위치와 크기를 결정하는 방법이다[12][13][14].

2.3 연구범위

본 논문에서는 대화형 방송 환경에서 부가서비스를 제공받기 위해 메뉴 검색을 하는 것이 아니라 방송되고 있는 영상의 화면 내부에서 원하는 객체를 선택했을 때 선택한 객체에 대한 부가서비스를 제공받는 새로운 방법을 제안하고 시스템을 설계하는 것이다. 이를 위해서는 디지털방송에서 사용되는 MPEG 멀티미디어 스트림 파일을 편집 가능한 프레임 단위로 분할하고 부가서비스 제공 대상인 객체를 선택하여 선택한 객체를 추적하는 기술이 필요하다. 또한 방송화면과 부가서비스 제공 대상인 객체정보(위치, 크기, 모양)의 동기를 맞추는 기술이 필수적이다. 따라서 본 논문의 연구범위는 마이크로소프트(Microsoft)에서 제공하는 다이렉트쇼(Direct Show)를 기반으로 멀티미디어 스트림 파일을 편집 가능한 프레임으로 나누고 각각의 프레임에서 선택한 객체를 추적하여 부가서비스 제공 대상인 객체정보(위치, 크기, 모양)를 생성하는 시스템을 구현하는 것이다.

3. 시스템구성

본 논문은 대화형 방송환경에서 부가서비스 제공 대상인 객체정보(위치, 크기, 모양)를 생성하는 동영상 편집 스튜디오를 설계하는 것이다. 이를 위해서는 대용량의 멀티미디어 스트림 데이타를 편집 가능한 프레임으로 나누고 프레임에서 부가서비스 제공을 위한 객체를 선택하며, 선택한 객체를 각각의 프레임에서 추적하여 부가서비스 제공 대상인 객체정보(위치, 크기, 모양)를 생성한다. 또한 이렇게 생성된 객체정보와 동영상 스트림 데이타가 동일한 시간에 재현 되도록 동기를 맞추는 것이다.

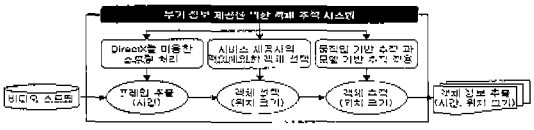


그림 3 시스템 구성도

▪ 프레임 추출

본 논문에서는 마이크로소프트사에서 제공하는 다이렉트쇼(Directshow)를 사용하여 멀티미디어 스트림 데이타를 프레임 단위로 분할하였고 프레임 동기를 맞추기 위한 시간정보는 프레임 추출 시간을 이용하였다[15].

▪ 객체 선택

부가서비스 제공대상인 객체 선택은 서비스를 제공하는 정보 제공자에 의해 선택되는데, 선택 방법은 크게 객체의 영역 선택과 객체 위치 선택으로 나누어진다. 추적하길 원하는 객체를 움직임 기반으로 추적하는 경우에는 영역 선택방법을 선택하고, 모델 기반으로 추적하는 경우에는 객체의 중심위치만 선택하면 된다.

▪ 객체 추적

비디오 데이타에서 부가서비스 제공 대상이 되는 특정 객체를 추적하기 위한 방법은 크게 움직임 기반 추적 방법(Motion-based Tracking)과 모델 기반 추적 방법(Model-based Tracking) 두 가지 접근 방법을 사용하고 있다. 최근 정확한 추적을 위한 방법으로는 모델기반 추적 방법이 많이 사용되고 있으나 실제 비디오에서 부가정보 삽입을 위한 객체는 종류가 너무 많기 때문에 모든 객체에 대해 모델을 만들 수 없다. 따라서 본 시스템에서는 대상 객체를 크게 사람과 사물로 나누어, 사람의 얼굴은 모델을 만들어 추적하는 모델 기반 얼굴 추적 방법을 사용하고 나머지 사물에 대해서는 객체의 영역을 지정하여 영역을 추적하는 움직임 기반 추적 방법을 적용하였다.

4. 객체 추적

4장은 본 논문에서 사용된 객체 추적 방법들에 대하여 서술한다.

4.1 움직임 기반 객체 추적

움직임 기반 객체 추적은 이전 프레임에서 객체에 해당하는 특정 영역이 다음 프레임의 검색 영역 중 어느 위치로 움직였는지를 찾는 방법이다. 본 논문에서는 시스템 사용자가 객체에 해당하는 영역을 선택하고 동작 벡터를 찾는 블록 정합 방법을 사용하여 영역의 움직임을 추적하였다. 또한 기존의 블록 정합 방법들에 시간적 예측 방법을 적용하여 일반적인 블록 정합 방법들로는 추적 할 수 없었던 움직임이 큰 객체에 대해서도 정확히 추적할 수 있는 방법을 제안한다.

4.1.1 블록 정합 검색 알고리즘

움직임 기반 추적 방법은 현재의 프레임을 구성하고 있는 화소들이 다음 프레임에서 어디로 움직였는지를 평가하는 방법이다. 이 방법은 광류(Optical flow)라고 불리는 시간적 공간적 변화량에 기초를 두고 있으며 크게 광류 방정식(Optical flow equation methods)[Lucas 81][Horn 81][Nagel 86]과 Pel-recursive methods [Walker 84] [[Biemond 87] 그리고 블록 정합(Block matching methods)으로 분류할 수 있다. 이중 블록 정합 방법은 코딩의 효율성과 적은 계산량 때문에 H.261/262/263그리고 MPEG-1/2와 같은 비디오 압축 방법에서 가장 많이 사용되는 움직임 평가 방법이다. 블록 정합 알고리즘은 그림4처럼 현재의 프레임(영상)을 여러 개의 블록(사각영역)으로 나누고, 각각의 블록이 다음 프레임의 검색 영역 중에서 어느 위치로 움직였는지를 판단하는 것이다. 블록의 크기를 가로세로 16×16화소 크기로 나누고 검색 영역 크기를 ±s로 설정하였다면 각각의 블록(사각영역)비교 지점 개수는 (2s+1)(2s+1)이고 각각의 비교 지점에서 다시 블록의 화소만큼(16×16) 비교하여야한다.

▪ 전역 검색(Full search : FS)

전역 검색 알고리즘은 검색 영역에서 움직일 수 있는

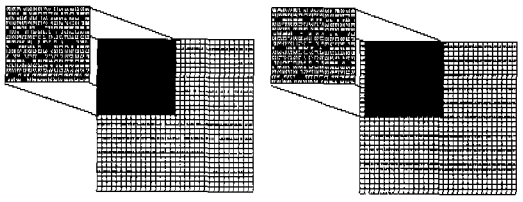


그림 4 블록 정합

모든 후보 지점을 비교하는 방식이다. 예를 들어 검색 영역 크기가 ±7이면 비교하여야 할 지점의 개수는 15*15(255)개 지점이다.

- 3단계 검색(Three-step search : 3SS)

3단계 검색 알고리즘은 검색 후보 지점의 개수를 대수적으로 줄이는 방법이다[17].

- 4단계 검색(Four-step search : 4SS)

4단계 검색 알고리즘은 검색 후보 지점 중 중첩되는 지점과 움직임의 크기를 판별해서 검색 지점의 개수를 줄이는 방법이다[18].

- 새로운 3단계 검색(New three-step search : N3SS)

새로운 3단계 검색 알고리즘은 3SS를 변형한 방법으로 대부분의 비디오 데이타의 움직임 변화가 ±3이내에서 움직인다는 특성을 고려한 방법이다[19].

- 유사도 측정(Matching criteria)

현재 프레임의 블록과 다음 프레임의 후보 영역 사이의 유사도를 측정하기 위한 방법은 Mean squar error(MSE)와 Mean Absolute error(MAE), Maximum matching pel count(MPC)를 사용하는데, 최근에는 하드웨어의 계산량이 가장 적은 MAE를 가장 많이 사용한다.

$$MAE(k,l;u,v)=\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} | I_n(k+i,l+j)-I_{n-1}(k+i+u,L+j+v)| \quad (3)$$

(k, l) : 현재 프레임 블록의 왼쪽 위 좌표

(u, v) : 이전 프레임 블록의 왼쪽 위 좌표

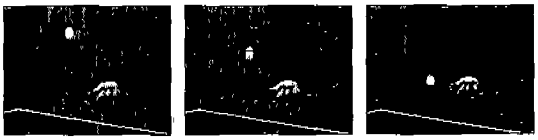
M, N : 블록의 가로, 세로 크기

4.1.2 제안된 시간적 예측 검색 알고리즘

4.1.1에서 살펴본 3SS, 4SS, N3SS방법들은 모두 공간적 예측(Spatial Prediction)의 특징을 이용한 추적 방법이다. 공간적 예측이란 현재 서울에서 비가 오고 있고 1시간 후에 어느 지역에 비가 올지 예상할 때 서울이나 또는 서울과 인접한 지역에 비가 올 가능성이 높다는 예측 방법이다. 이러한 공간적 예측을 이용하여 현재 프레임의 블록이 다음 프레임에서 어느 지점으로 움직였는지를 판별할 때 블록의 현재 위치와 그 주변 위치를 조사하여 가장 유사한 위치로 검색 위치를 이동시키고 이동시킨 위치에서 다시 주변 위치를 검색하는 방법이다.

공간적 예측과 함께 생각할 수 있는 방법은 시간적 예측(Temporal Prediction)인데 이것은 1시간 전에 의정부에서 비가 왔고 현재는 서울에서 비가 온다면 1시간 후에는 수원에서 비가 올 가능성이 높다는 의미이다. 이러한 시간적 예측을 이용하여 현재 프레임의 블록이 다음 프레임에서 어느 곳으로 움직였는지를 판별할 때 이전 프레임들로부터 움직인 방향과 움직인 거리의 변화량을 측정하여 다음 프레임에서의 객체 움직임 방향과 거리를 예측하고, 다음 프레임의 검색 위치를 예측된 위치로 이동하여 검색하는 방법이다. 이와 같은 시간적 예측은 객체 이동경로에서 많이 나타나는데 이것은 이론적인 특성이 아니라 관측된 실험적인 특성이다.

일반적인 블록 정합 검색 방법에서 성능테스트를 위해 사용하는 고전적 실험데이타들은 대부분 객체 움직임이 ±5화소 이내에 속하는 데이타들이다. 그러나 방송에서 사용되는 비디오 데이타에서는 ±5화소 이상 움직이는 객체도 많기 때문에 이렇게 장면진행이 급격하게 변하는 비디오 데이타에서는 객체 위치를 놓치는 경우가 많이 발생한다[그림 5 참조].



frame#7 frame#9 frame#11

그림 5 검색 영역 크기가 ±14인 일반적 전역검색 방법

그림 5는 검색 영역 크기가 ±14화소인 전역 검색 알고리즘으로 탁구공의 위치를 추적하는 그림이다. 그림에서 보는 것처럼 탁구공의 움직임이 큰 경우에는 검색 영역 크기가 ±14인 전역 검색 알고리즘으로도 정확한 추적을 할 수 없다. 일반적인 전역 검색 방법으로 탁구공을 추적하기 위해서는 검색 영역 크기를 22 이상으로 설정하여야만 추적에 성공을 할 수 있었다. 이러한 문제점을 극복하기 위하여 검색 영역 크기를 확대할 수 도 있지만 검색 영역 크기 확장은 곧 계산량의 증가로 이어져서 전체 시스템의 수행속도에 많은 영향을 미치게 된다. 따라서 본 시스템에서는 시간적 예측을 적용하여 검색 영역 크기 확장 없이 ±7화소 이상의 변화까지 해결하였다.

시간적 예측 검색 알고리즘(Temporal search algorithm)은 이전 프레임들의 블록 위치 변화량으로부터 다음 프레임에서 블록위치를 예측하고 다음 프레임의 검색 영역을 예측된 위치로 이동하여 검색하는 방법이다. 즉 객체 영역 중심좌표가 (7,7),(13,13),(20,20),(28,28)의 순서로 이동하고 있다면 다음 중심 위치는 (37,37)이 될 것으로 예측하여 검색 영역 중심위치를 (28,28)에서 ±7화소만큼 검색하는 것이 아니라 검색 영역 중심위치를(37,37)로 이동

하여 ±7화소를 검색한다. 이러한 시간적 예측 방법을 이용하면 객체 움직임이 ±7화소 이상 발생하여도 효과적으로 추적 할 수 있다. 다음 수식 4.1.2는 이러한 중심 좌표의 변화 예측을 간단히 모델링 한 것이다.

$$x_t^p = 3x_{t-1} - 3x_{t-2} + x_{t-3}$$
$$y_t^p = 3y_{t-1} - 3y_{t-2} + y_{t-3} \quad (4)$$

간단한 수식4.1.2의 예측만으로도 비디오에서 객체의 정지, 가속운동, 등속운동, 감속운동을 효과적으로 예상하여 검색 영역 크기 확장 없이 변화가 큰 움직임도 정확히 추적 할 수 있다. 객체의 운동이 (7,7),(7,7),(7,7)로 정지상태 때는 다음 프레임의 위치도 (7,7)로 예측되며, 객체의 운동이 (7,7),(10,10),(13,13)으로 등속상태 때는 다음 프레임의 위치는 (16,16)으로 예측된다 또한 (28,28),(20,20),(13,13)으로 감속상태 때는 다음 프레임의 위치는 (7,7)로 예측된다.

그림 6에서는 검색 영역 크기가 ±7인 전역 검색 방법이지만 이전 프레임들의 위치 변화량으로부터 다음 프레임의 위치를 예측하고 예측된 위치의 인접지역 ±7화소 이내를 검색하기 때문에 검색 영역 크기 확장 없이도 효과적인 추적이 이루어졌다. 이러한 과거 데이타로부터 미래를 예측하는 기법은 많이 있지만 본 시스템에서는 계산량을 줄이는 것이 목적이기 때문에 많은 계산량이 소요되는 미래 예측 방법들[12][13]은 검색을 위한 계산량 이외에 부가적으로 많은 계산이 사용되므로 검색 영역 크기를 확장하는 것만큼이나 계산량을 증가시킨다. 따라서 본 시스템에서는 부가적인 계산량을 최소화 할 수 있는 간단한 방법을 적용하였다.

본 시스템에서 제안하는 방법의 계산량을 살펴보면 현재 프레임의 블록 위치에서 다음 프레임의 블록위치를 결정하기 위한 블록 정합 계산량은 57600번(16×16(블록크기)×15×15(비교지점개수)뺄셈) 연산이 필요하며 블록 위치를 예측하기 위한 계산량은 16번(곱셈4, 덧셈2, 뺄셈2) 연산만 필요하다. 때문에 전체 블록 정합 연산과 비교하면 블록위치 예측 연산은 1/3600 수준이다. 따라서 전체 시스템의 속도에 미치는 영향은 무시할 정도로 극히 미미하다.

frame#7 frame#9 frame#11

그림 6 검색 영역 크기가 ±7이고 시간적 예측을 적용한 전역검색 방법

4.2 모델 기반 객체 추적

모델 기반 객체 추적방법에서는 얼굴 모델을 생성하고 각 프레임에서 독립적으로 얼굴을 찾아내는 방법으로, 본 논문에서는 색상 모델과 2-D타원 모델[16]을 이용하였다. 색상 모델을 이용한 방법은 조명, 배경에 따라서 많은 오류를 나타내고 얼굴이 회전하여 뒷모습이 나타날 때는 계속적으로 추적할 수 없다는 단점이 있고 2-D타원 모델을 이용한 방법은 영상에서 모양정보를 얻기 위한 이진화 영상 생성에서 정확한 에지 성분을 생성하지 못하거나 타원이외에 복잡한 에지 성분이 검출되면 역시 추적에 실패하는 단점이 있다. 따라서 본 논문에서는 색상 모델과 2-D모델 각각의 단점을 서로 보완하기 위해서 두 가지 모델을 동시에 적용하였다.

각각의 새로운 영상에서 얼굴 사이즈와 위치는 얼굴 모델에 의해서 결정된다. 얼굴 모델은 단축과 장축의 비가 1:1.3인 타원 모델과 타원의 내부 색상이 사용자에 의해서 선택된 얼굴색상과 동일한지를 조사하는 색상 모델을 결합하여 생성한다.

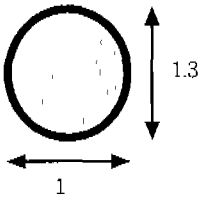


그림 7 타원 모델과 색상 모델을 결합한 얼굴 모델

$$S^* = \arg\max_{s_i \in S}\{\text{Ellipse\_precision}(s_i) + \lambda \ \text{Color\_precision}(s_i)\} \quad (5)$$

s_i = : (x, y, δ) 타원의 중심좌표와 단축길이

S : 검색영역

λ : 타원모델에 대한 색상모델의 가중치($\lambda \approx 0.3$)

수식 4.2는 검색 영역(S)에 포함되는 모든 타원 모델(s_i) 중 타원 정확도와 색상 정확도의 합이 가장 큰 값을 나타내는 타원 모델(s_i)을 얼굴(S^*)로 결정하는 것이다. λ는 타원 모델에 대한 색상 모델의 가중치로써 본 시스템에서는 0.3으로 결정하였다. 가중치를 1 이상으로 설정하면 타원 모델보다 색상 모델에 더 많은 영향을 받아서 얼굴이 회전할 경우에는 추적에 실패하는 경우가 발생하기 때문에 얼굴이 회전하거나 크기가 변형되어도 추적 할 수 있도록 하기 위해서는 λ를 1보

다 작은 값으로 선택하는 것이 합당하다. λ에 대한 값은 0.1부터 0.9까지 0.1씩 변화시켜본 결과 중 얼굴의 회전에 견고한 특성을 나타내는 0.3으로 선택하였다.

4.2.1 타원 모델

타원 모델을 이용한 방법은 단축의 길이와 장축의 비가 1 : 1.3인 2-D타원 모델을 만들고 원 영상에서 에지 정보를 추출한 이진화 영상을 생성하여 이진화 영상에서 타원 모델과 가장 잘 정합 되는 타원 모델의 위치와 크기를 찾아내는 것이다[16]. 타원 모델이 가장 잘 정합된다는 것은 타원 모델을 구성하고 있는 타원둘레 전체 화소 개수 중에서 몇 개의 화소가 이진화 영상과 교차시켰을 때 정합 되는지를 살펴보고 그 개수가 가장 많은 타원 모델(위치, 크기)이 가장 잘 정합되는 타원 모델로 결정하는 것이다. 즉 단축의 크기가 30화소인 타원 모델부터 20화소인 타원 모델을 이진화 영상의 검색영역에 모두 매핑 시켜보고 가장 잘 일치하는 타원을 얼굴영역으로 판단하는 것이다. 여기서 타원 모델을 형성하고 있는 픽셀의 위치에 이진화 영상의 화소가 에지 성분이면 $G_s(i)$의 값이 1씩 증가한다[그림8 참조].

그림 8 타원모델을 이용한 얼굴 추적 과정

$$\text{Ellipse\_precision} = \frac{1}{N_\delta} \sum_{i=0}^{N_\delta} | G_s(i) | \qquad (6)$$

N_δ : 단축의 크기가 δ인 타원 둘레를 형성하고 있는 화소의 개수

$G_s(i)$: $S(x, y, \delta)$에서 이진화 영상과 타원 모델이 정합 되는 화소의 개수

4.2.2 색상 모델

색상 모델을 이용한 방법은 시스템 사용자가 추적하길 원하는 얼굴을 선택할 때 선택한 영역의 색상을 모델로 생성하여 검색영역에서 가장 잘 정합되는 영역을 찾는 방법이다. 첫 번째 단계에서 영상을 RGB색상에서 HSV색상으로 변환한다. 이렇게 RGB 색상을 HSV색상으로 변환하면 다양한 조명환경에 더욱 견고한 특징을 나타낸다[Gary 98]. 두 번째 단계에서는 사용자가 선택한 영역의 휘도(Hue)와 채도(Saturation) 각각의 최대값과 최소값을 구하여 임계값을 설정한다. 이렇게 생성된 색상 모델을 검색 영역에 포함되는 모든 타원 모델 내부색상과 정합 하여 타원 모델 내부색상이 임계값 이내이면 얼굴 영역으로 판단하는 것이다. 색상 정확도는 타원내부 색상이 임계값이내에 들어오는 화소의 개수를 타원내부를 형성하고 있는 전체 화소의 개수로 나누어 준 값이다[20].

$$S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)]$$

$$H = Cos^{-1} \left[\frac{\frac{1}{2}[(R-G)+(R-G)]}{\sqrt{(R-G)^2+(R-B)(G-B)}} \right] \qquad (7)$$

H: 휘도(Hue), S: 채도(Saturation)

$$\text{Color\_precision} = \frac{1}{IN_\delta} \sum_{i=0}^{IN_\delta} | FR_s(i) | \qquad (8)$$

$FR_s(i)$: $S(x, y, \delta)$에서 타원의 내부 색상

$FR_s(i) = 1$, IF $T_{H1} \le H \le T_{H2} \bigcap T_{S1} \le S \le T_{S2}$

$= 0$, otherwise

IN_δ : 크기가 델타인 타원의 내부의 화소 개수

T_{H1}, T_{S1} : 사용자가 선택한 영역의 휘도 최소, 최대값

T_{S1} T_{S2} : 사용자가 선택한 영역의 채도 최소, 최대값

2-D타원 모델과 사용자가 선택한 영역의 색상 모델을 결합한 얼굴 추적 방법은 움직임은 작지만 움직임 보정(Motion compensate)방법에서 많이 사용되었던 실험 데이타에 대해서 실험을 하였다[그림 9 참조].

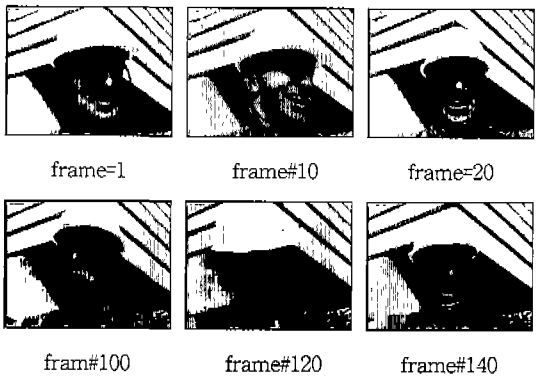


그림 9 Foreman에 대한 얼굴 추적 과정

5. 결 과

5.1 실험 데이타와 실험 환경

실험 데이타는 비교적 움직임이 많은 탁구(Tennis)영상을 대상으로 실험하였고 얼굴 추적방법을 실험하기

위해서는 얼굴의 회전과 다른 객체에 의한 얼굴의 가림(Occulution)현상을 인위적으로 발생시킨 데이타를 사용하였다. 또한 대화형 방송에서 사용될 일반적인 방송용 실험 데이타는 시스템이 사용되는 범주를 고려하여 실험데이타를 4개의 분야로 나누어 골고루 수집하여 실험하였다[Appendix참조]. 또한 실험은 다음과 같은 환경에서 이루어졌다.

CPU : Pentium-III 450, Memory : 256 MB,

개발언어 : Visual C++

5.1.1 시간적 예측 검색 결과

4.1.2절에서 제안하는 시간적 예측 검색방법이 어느 정도 정확한 검색이 이루어지는지 객관적인 판단을 위해서는 사용자에 의해서 선택되는 특정 객체 추적이 아닌 영상 전체에 대한 블록 정합 방법을 적용하여 비교해야 한다. 그러나 고전적으로 사용되는 Miss America, Clair, Sales man, Foreman, Garden, 같은 동영상들은 움직임이 대부분 ±1~+4화소 이내에서 움직이기 때문에 검색 영역 크기가 ±7인 일반적인 블록 정합 방법에서는 시간적 예측 방법을 적용하여도 Mean Absolute Error를 줄일 수 없다. 따라서 본 논문에서 제안하는 방법을 실험하기 위하여 검색윈도우 크기를 ±4로 줄여서 탁구(Tennis)영상에 적용하였다. 또한 스탠포드대학 웹사이트에서 움직임이 큰 스키(ski)영상을 다운로드 하여 실험하였다. 표 3과 표 4는 탁구영상과 스키영상에 대한 Mean Absolute Error를 하여 실험한 결과이다.

표 3 Ski 영상의 Mean absolute error, 검색윈도우 ±7, frame#1~frame#60, TPS(Temporal Prediction Search)

Temporal prediction \ Block matching	FS	3SS	4SS	N33
without TPS	14.626	14.940	15.242	15.818
with TPS(제안된 방법)	12.443	12.844	12.900	12.896

표 4 Tennis 영상의 Mean absolute error, 검색 윈도우 ±4, frame#1~frame#60, TPS(Temporal Prediction Search)

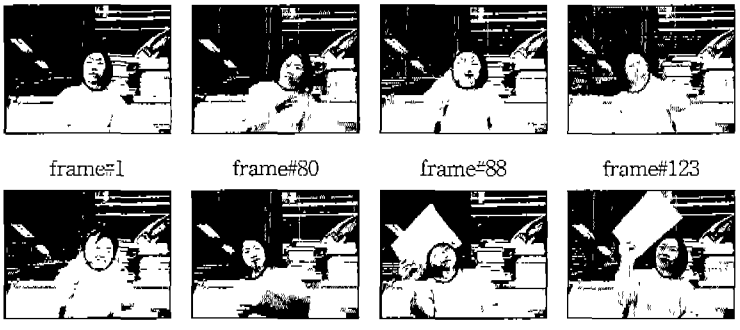
Temporal prediction \ Block matching	FS	3SS	4SS	N33
without TPS	6.923	7.337	7.190	7.030
with TPS(제안된 방법)	6.876	7.056	7.053	7.026

5.1.2 모델 기반 객체 추적 결과

4.2절에서 제안한 얼굴 추적방법이 방송용 비디오에서도 견고하다는 것을 실험하기 위해서 얼굴의 회전과 다

frame#1 frame#70 frame#100 frame#150

그림 10 회전(Rotation)



frame#1 frame#80 frame#88 frame#123

frame#154 frame#198 frame#205 frame#210

그림 11 기울임(tilt)과 가림(Occlusion)

른 객체에 의한 얼굴의 가림(Occlusion), 기울임, 얼굴의 크기변화가 발생하는 비디오 데이타가 필요하며, 이를 위해 급격한 얼굴의 움직임과 회전, 얼굴 가림, 기울임, 줌 현상을 인위적으로 생성하여 실험하였다.

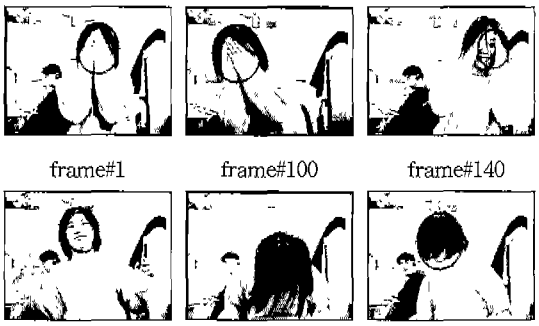


그림 12 기울임(tilt)과 줌(zooming)

6. 결론 및 향후 연구과제

본 논문에서 제안한 방법은 대화형 방송환경에서 시청자가 부가서비스를 제공받기 위해 메뉴검색을 하는 것이 아니라 영상에서 원하는 객체 선택만으로 부가서비스를 제공받는 방법에 대하여 제안하였다.

시청자가 원하는 객체 선택을 위해서는 방송영상과 동기를 맞춘 객체정보(위치, 크기, 모양)를 제공해야 해야하는데 동기를 맞추는 기술은 마이크로소프트(Microsoft)에서 제공하는 다이렉트쇼(DirectShow)를 이용하였고, 동영상에서 객체를 추적하기 위한 방법은 모델 기반 객체 추적과 움직임 기반 영역 추적 두 가지 방법을 사용하였다.

부가서비스를 제공하기 위한 객체는 크게 사람과 사물로 나누어지는데, 사람은 얼굴에 대한 타원 모델과 색상 모델을 만들어 각각의 프레임에서 모델과 가장 잘 정합되는 모델의 위치와 크기로 얼굴을 추적하였고 사물은 시간적 예측 검색 방법을 적용한 블록 정합 방법을 사용하여 추적하였다. 시간적 예측 검색 방법의 적용으로 블록 정합에서 검색영역크기 ±7화소만으로도 ±14화소 이상 움직이는 객체(영역)도 효과적으로 추적하였다.

그러나 지상파 방송에서 사용되는 영상들은 카메라의 줌인(Zoom in)이나 줌아웃(Zoom out)에 의하여 객체 크기가 변화거나 카메라의 위치 이동으로 객체 모양이 변화는 경우도 발생한다. 물론 얼굴 추적 방법에서는 이러한 줌(Zooming)이나 카메라의 위치 이동에도 견고한 시스템을 구성하였지만 일반적인 사물에 적용한 움직임 기반 추적 방법에서는 크기가 변화되거나 모양이 변화되는 객체에 대해서는 정확한 추적을 할 수 없다. 이러한 문제점을 정리해 보면 다음과 같다.

- 카메라의 줌(Zooming)에 의한 객체 크기변화
- 카메라의 기울임(Tilt)에 의한 객체 모양변화
- 카메라의 위치이동에 의한 객체 모양, 크기변화
- 객체의 회전에 의한 객체 모양변화
- 객체의 움직임에 의한 객체 모양, 크기변화

이러한 문제점을 해결하기 위해서는 움직임 기반 추적 방법에서 고정된 크기 움직임 추적이 아니라 가변적 크기 움직임 추적 방법이 필요하며, 이를 해결하기 위한 연구가 진행 중에 있고, 또한 객체 기반 코딩으로 많은 주목을 받고 있는 MPEG-4기반 객체 추적방법으로 해결하기 위한 노력도 진행 중에 있다. 이러한 연구가 성공적으로 끝나면 지금보다 더욱 정교하게 객체 추적을 할 수 있을 것이다.

마지막으로 대화형 방송환경에서 영상내부의 객체에 대한 효과적인 부가서비스 제공을 위해서는 방송기획단계부터 부가서비스 제공 객체와 광고모델을 고려한 방송제작이 필요하다.

참 고 문 헌

[1] http://www.atsc.org.
[2] http://www.dvb.org.
[3] www.opentv.com.
[4] 한국방송공사(KBS), "대화형 디지털방송[iPCTV] 시연회 발표집", 2코엑스 인터컨티넨탈 서울, pp. 30-31, 2000.
[5] http://www.microsoft.com/TV/ITVSAMPLES/TVsamples.htm
[6] D. Beymer p. McLauchlan, B. Coifman,and J. Malik. "A real-time computer vision system for measuring traffic parameters," In Proceedings of IEEE CVPR 97, page 495-501, 1997.
[7] K. Kanatani, "Camera rotation invariance of image characteristics," Computer Vision, Graphics and Image Processing, Vol. 39, No. 3, Sep. 1987, pp. 328-354.
[8] M.Hunke and A. Waibel, "face locating and tracking for human-computer interaction," Proceedings of the 28th Asilomar Conference. On Signals, Sys. and Comp., pp. 1277-1281, 1994.
[9] A. Azarbayejani, T. Darrell, A.Pentland, "Pfinder: Real-Time Tracking of the Human Body," SPIE Vol. 2615, 1995.
[10] K.Sobottka and I. Pitas, "segmentation and

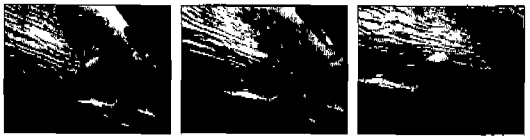
tracking of faces in color images," Proceeding Of the Second Intl. Conference On Auto. Face and Gesture Recognition, pp. 236-241, 1996.
[11] P.Fieguth and D. Terzopoulos, "Color-based tracking of heads and other mobile objects at video frame rates," In Proceedings of IEEE CVPR, pp. 21-27, 1997.
[12] A. Blake, R. Curwen, and A. Zisserman. "A framework for spatiotemporal control in the tracking of visual contours," Intl. Journal of Computer Vision, 11(2): 127-145, 1993.
[13] C. Le Buhan, F. Bossen, S.Bhattacharjee, F. Jordan, T. Ebrahimi, "Shape representation and coding of visual objects in multimedia applications - An Overview," (invited paper), in Annales des Telecommunications 53, No 5~6, pp. 164-178, May 1998.
[14] D. Murray and A. Basu, "Motion tracking with an active camera," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 16, No. 5, May 1994, pp. 449-459.
[15] http://www.micorsoft.com/directx.
[16] Stan Birchfield, "An Elliptical Head Tracker," 31st Asilomar Conference on Signals, Systems, and Computers, November 1997.
[17] T. Koga, K, Iinuma, A. Hirano, Y.Iijima, and T. Ishiguro, "Motion compensated interframe coding for video conferencing," Proceedings NTC81, pp. G5.3.1 5.3.5, Nov. 1981.
[18] L. M. Po and W.C.Ma, "A novel four-step search algorithm for fast block motion estimation," IEEE Transaction Circ. Syst. and Video Technology, vol. 6, pp.313-317, June 1996.
[19] R.Li, B. Zeng, and M. L. liou, "A new three-step search algorithm for block motion estimation," IEEE Transaction. Circ. Syst. and Video Technology, vol. 4, pp. 438-442, Aug. 1994.
[20] 고병철, "A Study on Content-Based Image Retrieval System Supporting Dynamic search environment," 연세대학교 대학원 석사졸업논문, 2000.

APPENDIX

frame#1 frame#10 frame#20

frame#30 frame#40 frame#50

Appendix 1 다큐멘터리(Documentary) 1



frame#1 frame#30 frame#70

Appendix 2 다큐멘터리(Documentary) 2

frame#1 frame#10 frame#20

Appendix 3 스포츠(Sport)

frame#1 frame#20 frame#50

frame#1 frame#30 frame#50

Appendix 4 오락(Entertainment)

frame#1 frame#100 frame#150

Appendix 5 드라마(Drama)

frame#1 frame#10 frame#20

Appendix 6 드라마(Drama)2

안 준 한

1999년 명지대학교 컴퓨터공학과 졸업(공학사). 2001년 연세대학교 대학원 컴퓨터과학과 졸업(공학석사). 2001년 ~ 현재 LG전자 DTV 연구소 연구원. 관심분야는 비디오 인덱싱, 객체추적

변 혜 란

1980년 연세대학교 수학과 졸업(이학사). 1983년 연세대학교 대학원 수학과 졸업(이학석사). 1987년 Univ. of Illinois, Computer Science(M.S.). 1993년 Purduc Univ., Computer Science(Ph.D.). 1994년 ~1995년 한림대학교 정보공학과 조교수. 1995년 ~ 1998년 연세대학교 컴퓨터과학과 조교수. 1998년 ~ 현재 연세대학교 컴퓨터과학과 부교수. 관심분야는 인공지능, 영상인식, 영상처리