

대화형 하이퍼 비디오 저작을 위한 객체 기반 추적 방법

황본우⁰, 손형진 / 이성환

(주)버추얼미디어 기술연구소⁰ / 고려대학교 인공시각연구센터

{bwhwang,hjson}@virtualmedia.co.kr, swlee@image.korea.ac.kr

An Object-based Tracking Scheme for Interactive HyperVideo

Bon-Woo Hwang⁰, Hyung-Jin Son / Seong-Whan Lee

Virtual Media, Inc.⁰ / Center for Artificial Vision Research, Korea University

요약

컴퓨터 비전 기술을 이용한 효과적인 객체 추적 기술은 인공 시각, 컴퓨터와 인간의 상호작용(HCI), 영상 기반 제어 장치, 감시 시스템 등 다양한 분야의 응용에 있어 중요한 연구 과제이다. 특히 최근 들어 초고속 통신망의 보급으로 인해 인터넷 방송과 같은 실시간 동영상 전송 기술이 가능해짐에 따라 단순한 텍스트와 정지영상상을 제공하는 하이퍼텍스트 환경에서 사운드를 포함하는 동영상 데이터를 제공하는 하이퍼미디어 환경으로 변하고 있다. 이러한 하이퍼미디어 환경에서의 객체 추적은 객체 단위의 링크가 가능한 한 하이퍼비디오 구현에 있어서 필수적이다. 따라서 본 논문에서는 다양한 동영상 데이터 상의 객체를 효과적으로 추적하는 알고리즘을 제안하고자 한다.

1. 서 론

인간의 시각 시스템과 같이 실시간으로 객체를 추적하는 것을 컴퓨터 비전 기술을 이용하여 구현하기란 매우 어려운 일이다. 하나의 프레임이 많은 제약 조건하에서 영상으로 구현되기 때문에 영상 자체만으로는 그 장면을 분석하기에 충분한 정보가 제공되지 않으며, 실제세계의 3차원 장면이 2차원 영상으로 투영될 때 깊이에 관한 정보가 붕괴되기도 한다.

이러한 어려움으로 인해 컴퓨터 비전 시스템을 이용한 인공 시각, 컴퓨터와 인간의 상호작용, 영상 기반 제어 장치, 감시 시스템 분야에서 사용되는 객체 추적 방법은 각기 특정 환경에서 잘 동작하도록 몇 가지 제약 조건을 두어 설계되고 있다. 카메라의 동작을 제한하여 객체와 배경의 분리를 용이하게 하는 방법, 추적하고자 하는 객체를 미리 정하여 그 객체의 특징과 형태를 미리 학습시켜 추적하는 방법, 또는 추적하고자 하는 객체의 수를 제한하는 방법 등이 그것이다. 이러한 방법으로는 본 연구에서 해결하고자 하는 다양한 동영상에서 나타나는 임의의 객체를 추적하는 방법에 대해서는 적합하지 않다. 카메라의 움직임이 고정되지 않은 상태에서, 나타나는 배경, 조명, 객체의 색, 모양 등이 다양한 객체를 추적하기 위해서는 다른 방법이 필요하다.

이에 본 논문에서는 일반적인 동영상 데이터 상에서 나타나는 임의의 객체를 효과적으로 추적하는 방법을 개발하고 이를 직접 그 소프트웨어적으로 구현하는 알고리즘을 제안하고자 한다.

본 연구 제안하는 방법은 우선 DirectShow를 기반으로 하여 일반 동영상으로부터 영상을 획득한다. 그 다음 획득한 영상을 기반으로 객체 추적의 시점과 종점을 결정

하기 위하여 장면 전환된 프레임을 검출한다. 그리고 검출된 프레임에서 추적하고자 하는 객체를 추출한다. 이렇게 추출된 객체를 다음 장면 전환이 발생하기 전 프레임까지 효과적으로 추적한다.

2. 관련 연구

추적할 객체가 미리 정해져 있지 않은 상태에서의 객체 추적은 어떠한 객체를 추적할지 선택하는 단계가 필요하며 이러한 선택 이후에 객체 추적이 시작되게 된다.

전체 영상에서의 광류 정보를 이용하여 영상에서의 객체와 배경을 분할한 뒤 분할된 객체를 추적하는 방법을 사용할 수 있다. Smith[5]의 경우 이러한 독립적 광류(Independent optical flow)를 이용하여 배경과 객체를 분할한 뒤, 객체 주변에서 나타나는 광류들의 방향성을 사용하여 객체의 움직임을 예측하였다. 이와 같은 방법은 배경의 변화나 카메라의 움직임이 비교적 적은 환경 하에서는 효과적으로 잘 적용될 수 있으나 추적할 객체의 움직임이나 카메라의 움직임이 큰 경우 또는 객체가 크게 클로우즈-업 되어있는 경우의 객체 추적 방법으로는 적합하지 못하다. Leymarie[5]는 시점에 따라 형태가 변하는 상황을 고려하여 객체를 추적하는 방법을 제안하였다. 객체를 둘러싼 변형 모델, Active Contour Model(ACM)을 사용하여 객체의 윤곽선으로 일치시키고 객체의 윤곽선을 추출하였다. 이 방법의 경우 배경이 단순한 경우에는 효과를 거둘 수 있으나 애지가 많은 배경 위에 추적 객체가 존재하는 경우 등 예지 오검출이 많이 일어나는 상황에서는 부적합하다.

3. 객체 추적 알고리즘

동영상에서 객체를 효과적으로 추적하기 위하여 객체의 애지 부분의 특징 벡터를 추출하여 그 특징 벡터의 광류를 추출한다. 다음으로 줌-인/줌-아웃 정도를 판단하

기 위하여 각 방향별로 분류된 광류가 줌-인/줌-아웃 성질을 갖는지 검사하고 정확한 줌-인/줌-아웃 크기를 판단하기 위하여 기학습된 광류 크기별 학습 모듈에서 가장 유사하다고 판단된 크기로 형판을 생성하게 된다. 이를 그림 1과 같이 순서도를 나타내었다.

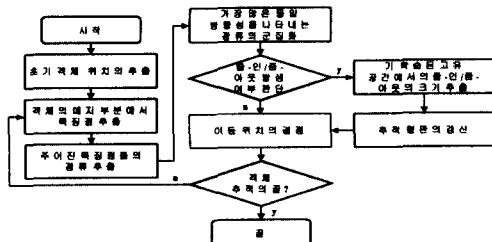


그림 1 제안된 객체 추출 방법의 블록 다이어그램

3.1 추적 객체의 특징점 추출

객체를 추적하기 위하여 객체를 인식하는데 있어서 중요한 특징 중의 하나인 에지 정보를 사용한다. 본 논문에서는 영상 전체의 화소를 특징점으로 사용하는 대신, 선택된 그림 2와 같이 객체 영상에서 에지를를 Canny 알고리즘을 사용하여 추출하고 추출된 에지 영역의 화소를 포함하는 $m \times m$ 의 블럭을 하나의 특징점으로하여 객체를 추적한다.

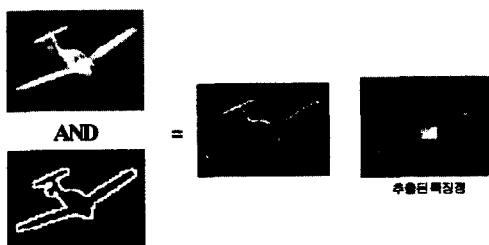


그림 2 특징점 추출 과정

3.2 선택적 광류를 이용한 객체의 이동 위치 검출

3.1절에서 언급한 방법으로 추출된 $m \times m$ 크기 벡터들의 다음 프레임에서의 위치를 최소차 알고리즘을 사용하여 검출하는 방법으로 각 특징점들의 광류 모델을 생성하고, 각 특징점들의 다음 프레임에서의 위치를 검출하는 식은 아래 식 (1)과 같다.

$$\sum_{(u,v) \in W} (I_1(u,v) - I'_1) - (I_2(x+u,y+v) - I'_2) \quad (1)$$

식 (1)에서 I_1 과 I_2 는 특징 벡터 원소, I'_1 와 I'_2 는 한 프레임의 특징 벡터들의 평균을 의미한다.

3.3 형판 생성을 위한 줌-인/줌-아웃 벡터 크기 학습

추적 과정에서 줌-인/줌-아웃이 일어났을 경우 추적 객체의 형판 생성을 위하여 학습된 고유공간에서의 줌-인/줌-아웃 크기와의 유사도 비교하여 형판생성을 한다.

이를 통해 객체의 줌-인/줌-아웃 발생시 형판 생성을 할 때 일어날 수 있는 여러 누적 현상을 방지 하였으며, 객체의 줌-인/줌-아웃이 일어날 때 나타나는 각 방향으로의 벡터 패턴을 그림 3과 같이 나타내었다.

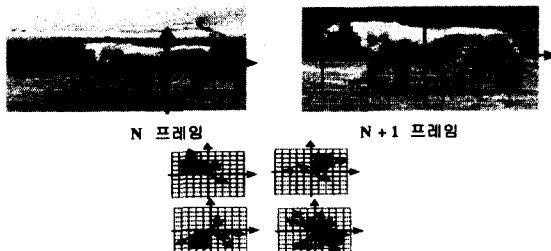


그림 3 줌-인 시의 각 방향별로 나타나는 광류들의 패턴

고유 벡터는 학습 영상의 공분산 행렬 K-L transform으로 계산할 수 있고 고유 벡터는 이 공분산 행렬로부터 얻어진다. 학습 영상의 고유 벡터를 추출하기 위하여 SVD(Singular Value Decomposition)을 사용하였다. 이 과정을 그림 4와 같이 나타내었다.

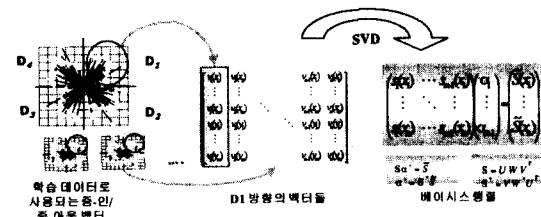


그림 4 벡터들의 패턴을 고유 공간으로 매핑시키는 과정

상기의 베이스 행렬을 통해 줌-인/줌-아웃이 발생하였을 때의 영상을 고유공간으로 매핑하여 이러한 광류들이 고유공간에서 어떠한 패턴을 나타내는지 각 패턴들의 파라메터 (a)를 구한다.

$$\begin{bmatrix} s_{11}(x_1) & \dots & s_{m-1}(x_1) \\ \vdots & \ddots & \vdots \\ s_{11}(x_n) & \dots & s_{m-1}(x_n) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_{m-1} \end{bmatrix} = \begin{bmatrix} S'(x_1) \\ \vdots \\ S'(x_n) \end{bmatrix} \quad (2)$$

$$a' = S' \cdot S' \quad (3)$$

여기서 S 는 정방 행렬이 아니므로 역행렬을 구할 수 없다. 이를 SVD(Singular Value Decomposition)을 이용하여 S 의 역행렬을 구할 수 있다.

$$S = U \cdot W \cdot V^T \quad (4)$$

$$S' = V \cdot W' \cdot U^T \quad (5)$$

$$a' = V \cdot W^T \cdot U_T \cdot S' \quad (6)$$

U의 열 벡터는 $S S^T$ 의 고유 벡터, V의 열 벡터는 $S^T S$ 의 고유 벡터이다. W 대각 원소들은 이 두 행렬의 제곱근으로 얻어진다. W' 는 W의 주요 대각 원소 중 0이 아닌 반비 수로 구성된다.

3.4 추적 객체의 특징 정보 생성

3.3절에서 학습한 줌-인/줌-아웃 크기의 패턴과 발생한 줌-인/줌-아웃의 패턴간의 유사도 비교를 통하여 오차 발생률을 줄일 수 있다.

4. 실험 결과 및 분석

4.1 실험 환경

본 논문에서는 마이크로소프트사의 윈도우 2000 환경을 기반으로 Visual C++ 6.0을 사용하여 구현하였다.

운영체제	MS Windows2000 Professional
CPU	Pentium III 800 MHz
RAM	128MB
Library	Intel Image Processing Lib. Open Source Computer Vision Lib.

표 1 실험 환경

4.2 결과 영상

본 연구에 사용된 데이터는 MPEG-1/2, AVI, ASF 등 다양한 형식의 동영상 데이터를 사용하였다.

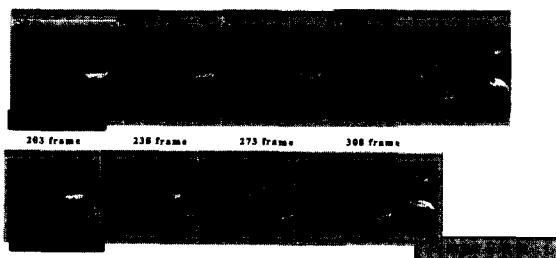


그림 5 추적 결과의 캡쳐 화면 I

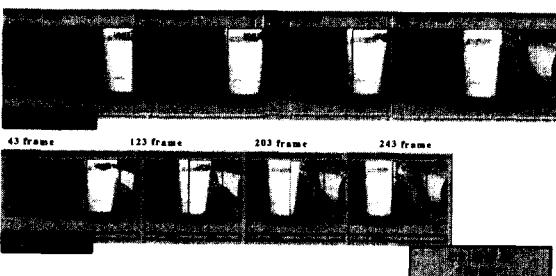


그림 6 추적 결과의 캡쳐 화면 II

4.3 실험 분석 및 향후 연구 방향

동영상에서 나타나는 객체들 중 임의로 선택된 객체에

대하여 추적된 결과를 Veon Studio와 비교하였다.

그림 5와 같이 객체의 급격한 움직임 일어난 경우 Veon Studio는 객체의 움직임을 다 따라가지 못하는 현상이 발생하게 된다. 이는 객체의 움직임으로 일어난 주된 광류의 방향보다 노이즈에서 발생한 광류의 움직임이 객체의 움직임 추적에 영향을 주기 때문이다. 그러나 본 논문이 제안한 방법으로 추적을 한 경우, 객체의 특징 이외의 잡음이 발생시킨 광류를 제거하고 추적함으로 안정적으로 객체를 추적함을 볼 수 있다.

그림 6과 같이 객체와 배경의 특징 정보가 크게 나타나지 않을 경우, 객체 특징점의 광류가 올바로 추출되지 않는다. 그러나 본 논문의 경우에는 검출된 객체의 광류들을 주된 방향성을 나타내는 광류들 이외에는 제거하여 객체 움직임 방향을 예측하므로 무리없이 객체의 추적이 가능함을 볼 수 있다. 본 연구에서 제안한 방법은 추적 할 객체의 특징에 대한 사전 학습 과정 없이 주요 특징 벡터를 추출하여 이를 검출함으로서 실시간에 가깝게(80 ms/1 frame) 객체를 추적할 수 있다. 하지만 초기 추적 객체의 특징과 추적 중에 나타난 객체의 특징이 심하게 달라질 경우엔 본 논문에서 제시한 추적 방법으로도 객체 추적에 실패하게 된다. 이를 개선하기 위해서는 객체의 이동 경로를 기억하여 템포 차의 크기를 능동적으로 조절하고 객체의 회전 변환, 가리워짐 현상에도 안정적으로 객체를 추적하여 추적 객체의 형판을 생성하여줄 수 있는 알고리즘을 추가한다면 보다 좋은 성능을 기대할 수 있을 것이다.

참고 문헌

- [1] K. Sobottka and I. Pitas, "Segmentation and Tracking of Faces in Color Images", Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Killington, Vermont pp. 236-241, 1996.
- [2] J. Yang and A. Waibel, "A Real-Time Face Tracker", IEEE Workshop on Applications of Computer Vision, Sarasota, FL pp. 142-147, 1996.
- [3] A. Colmenarez, B. Frey and T. S. Huang "Detection and Tracking of Faces and Facial Features", IEEE International Conference on Image Processing, Vol. 4, Kobe, Japan pp. 657-661, 1999.
- [4] F. Oberti and C. Regazzoni, "Adaptive Tracking of Multiple Non Rigid Objects in Cluttered Scenes", Proceedings of International Conference on Pattern Recognition, Barcelona, Spain pp. 1108-1112, 2000.
- [5] S. M. Smith, "ASSET-2: Real-Time Moti Segmentation and Object Tracking", Real-Tim Imaging, Vol. 4, pp. 21-40, 1998.
- [6] F. Leymarie and M. D. Levine, "Tracking Deformable Objects in the Plage using an Active Contour Model", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 6, pp. 617-634, 1996.