영상에서의 얼굴인식

한준희*, 남기환*, 정연숙*, 정주병***, 나상동**, 배철수*

*관동대학교, **조선대학교, ***(주)KT

A Study on Recognizing faces in broadcast video.

Jun-hee Han*, Kee-hwan Nam*, Youn-sook Joung*, Joo-byeong Jeong***, Sang-dong Ra**, Cheol-soo Bae*

*Kwandong University, **Chosun University, ***KT

E-mail : hanjh123@empas.com

요약

최근 영상 자료의 저장과 검색을 위한 시스템이 많이 연구되고 있다. 방대한 양의 영상 자료를 디지털화하여 파일로 저장하고 영상에 관한 각종 정보를 데이터베이스로 구성한 뒤, 키워드 등을 사용하여 필요한 영상을 네트워크를 통하여 검색하고 이것을 편집 등에 활용할 수 있도록 하는 것이 본 논문의 목적이다. 영상을 데이터베이스로 구축하기 위해 선행되어야 할 것은 연속적인 장면마다 또는 의미 있는 장면마다 영상을 분류하는 작업이다. 본 논문에서는 MPEG 비트스트림을 분석하여 장면 전환 지점을 자동으로 찾는 실험을 워크스테이션을 통하여 시행하였으며 기존 실행한 실험을 바탕으로 PC상에서 동영상 검색 시스템을 구현하였다. 동영상 검색 시스템은 뉴스, 드라마는 물론 각종 보안 영상 등 다양한 분야의 영상을 분석하여 장면 전환 지점을 찾고, 각 장면의 대표 영상을 저장한 뒤, 네트워크 환경에서 동영상을 검색할 수 있도록 만든 시스템이다.

I. 서론

최근 들어 얼굴인식에 대한 관심이 증가하여 다양한 분야들, 특히 방송 및 보안 분야에 급속히 적용되고 있는 시점이다. 그러기에 본 논문에서는 비디오에 있는 얼굴들을 인덱싱하여 화자를 구별할 목적으로, 비디오 인덱싱을 위한 얼굴인식 시스템을 제안하고자 한다.

얼굴 인식에 있어 가장 먼저 수반되어야 할 문제는 영상에서 얼굴을 검출해내는 것이다. 얼굴 검출은 이동, 크기변화 및 회전에 독립적으로 인식 하도록 하여 주며, 얼굴 특징들의 위치에 대한 양질의 초기 조건들을 제공할 수 있다. 인식의 첫 번째 단계는 피부색조 픽셀들의 비율이 문턱값보다 더 큰지를 간단하게 결정하는 컬러 분할이다[1]. 이어서 Fisher Linear Discriminant(FLD)에 기초하여 후보 영역들 각각에 수치를 부여한다. 이 수치는 원래 많은 수의 얼굴 및 비얼굴 단편들을 비교함으로써 만들어진다[2]. 후보는 그들이 학습에서 사용되는 많은 수의 얼굴 영상들 중 하나와 어느 정도 비슷한가에 대한 측정인 Distance From Face Space(DFFS)상에서 점수가 매겨진다[3]. 조합된 문턱값을 초과하는 모든 후보 영역들은 두 얼굴들이 중복되지 않는다는 조건들을 적용한 후에, 얼굴들로서 다양한 장르의 영상을 분석하여 장면 전환 지점을 찾고, 각 장면의 대표 영상을 저장한 뒤, 네트워크 환경에서 동영상을 검색할 수 있도록 만든 시스템이다.

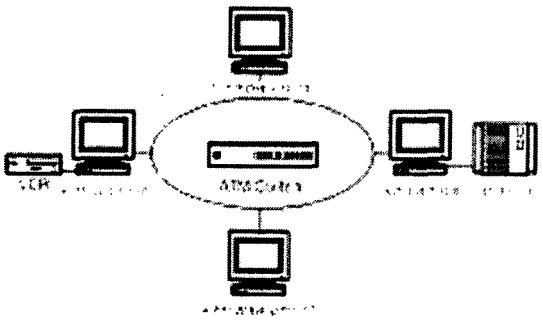

그림 1. 동영상 검색시스템의 구성

II. 얼굴영상 인덱싱 방법

동영상 인덱싱을 위한 첫 단계는 영상을 하나의 카메라에 의한 연속되는 장면단위로 나누는 것이다. 즉, 각각의 연속되는 장면이 전환되는 지점을 자동으로 찾는것이다[4]. 이렇게 검출된 각각의 연속적인 장면을 클립이라고 한다. 하나의 클립에 담긴 프레임들은 거의 비슷한 정보를 갖고 있으므로 동영상의 검색을 위한 최소 단위로 클립을 사용하게 된다. 각각의 클립을 가장 잘 표현하는 하나의 프레임을 대표영상이라 하고 보통 클립의 첫 프레임이 사용된다.

장면이 전환될 때 컷으로 전환되는 경우와 점진적으로 전환되는 경우가 있다.

2.1 컷 검출

컷 검출을 위한 대표적인 방법은 화소 비교법과 히스토그램 비교법이다. 화소비교법은 연속되는 프레임간에 대응하는 각각의 화소의 밝기의 차를 구하여 화면전체에서 얼마나 많은 밝기의 변화가 일어났는지를 계산하여 장면 전환 여부를 판단하는 방법이다.

$$DP_i = \sum_{k=0}^{M}\sum_{l=0}^{N}\left|Y_i(k,l) - Y_{i+1}(k,l)\right| \qquad (1)$$

DP_i : i 번째 프레임에서 밝기의 차

$Y_i(k,l)$: i 번째 프레임내의 픽셀 (k, l)의 휘도의 값

M : 수직 방향의 화면 픽셀 수

N : 수평 방향의 화면 픽셀 수

DP_i 가 특정 임계값보다 크면 장면 전환지점으로 판단한다. 그러나 이 방법은 카메라의 움직임에 민감한 단점이 있으므로, 움직임이 많은 영상에서는 컷 검출 오류를 일으킨다.

히스토그램 비교법은 화소의 세기(*Y* 성분)나 색상(*Cb*, *Cr* 성분)을 히스토그램으로 표현하여 식 (2)처럼 히스토그램의 차이로 두 인접 프레임간의 유사도를 측정한다. 이 방법은 화소 비교법에 비해 카메라나 물체의 움직임에 덜 민감한 장점이 있으나, 같은 물체를 다른 각도에서 촬영한 영상의 경우처럼 장면 전환이 일어났지

만 비슷한 히스토그램을 가진 인접한 영상의 경우에는 장면 검출이 불가능하다.

$$SD_i = \sum_{j=0}^{G}\left|H_i(j) - H_{i+1}(j)\right| \qquad (2)$$

SD_i : i 번째 프레임에서 히스토그램 차

G : 그레이 또는 컬러 레벨의 수

$H_i(j)$: i 번째 프레임에서 j 레벨에 해당하는 히스토그램 값

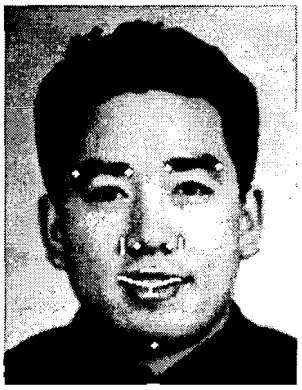

그림 2. 얼굴영상 특징점

본 논문에서 구현한 동영상 검색 시스템에서는 화소 비교법과 히스토그램 비교법을 결합하여 사용하였다.

즉, 먼저 히스토그램 비교법으로 컷 검출을 하고 히스토그램의 차이가 비교적 작은 경우에는 화소 비교법을 적용했다. 이러한 방법을 사용함으로써 히스토그램 비교법의 단점을 보완해 만족할만한 결과를 얻었다.

2.2 점진적 장면 전환 검출

컷으로 바뀌는 장면 전환의 검출은 비교적 간단하나, 특수 효과(디졸브, 페이드인, 페이드 아웃, 와이프)가 있는 장면 전환의 경우에는 장면 검출을 위한 다른 방법이 필요하다. 점진적 장면 전환의 경우에는 프레임간의 변화가 비교적 작아 단순한 화소 비교법이나 히스토그램 비교법만으로는 장면 검출이 불가능하다. 많이 쓰이는 이중 비교법(twin - comparison method)은 두 개의 임계치를 사용하여 점진적 장면 전환을 검출하는 방법이다. 두 개의 임계치는 각각 다음과 같다.

T_b : 컷 검출을 위한 높은 임계값

T_s : 특수 효과 검출을 위한 낮은 임계값

일단 히스토그램 차이값이 Ts 보다 크면 점진적 변화의 잠재적인 시작점(Fs)이라고 간주한다. 영상이 점진적으로 변하는 동안 연속되는 프레임과 잠재적인 시작점과의 히스토그램 차이값을 계산한다. 일반적으로 이 값은 시간적으로 점점 증가 하므로 누적 차이값(accumulated difference)이라 부른다. 만약 계산된 값이 Tb 를 초과하면 장면 전환 지점으로 결정한다. 이 값이 Tb 를 초과하기 전에 인접한 프레임간의 차이값이 Ts 보다 작아지면 그 잠재적인 시작점을 제거하고 다른 점진적 장면 전환 지점을 찾는다.

이중 비교법의 핵심은 두 개의 분명한 임계치 상태가 동시에 만족되어야 한다는 것이다. 이 방법의 단점은 연속적인 차이가 Ts 보다 작은 값을 가지는 동안에도 점진적인 장면 전환이 일어날 수 있다는 점이다. 이 문제를 해결하기 위한 방법으로 낮은 차이 값을 가진 연속적인 프레임의 수가 사용자가 허용하는 허용치 내에만 속한다면, 잠재적인 시작점을 제거하지 않음으로써 장면 전환 검출 오류를 방지한다.

2.3 검출 시간을 단축시키는 알고리듬

장면전환 지점을 검출할 때 걸리는 시간을 줄이기 위한 방법으로서 공간적 표본화와 시간적 표본화를 사용할 수 있다. 공간적 표본화는 영상을 공간적으로 표본화하여 실제보다 작은 크기의 영상을 취하여 계산하는 방법이고, 시간적 표본화는 한 프레임씩 이동시켜가며 비교하지 않고 여러 프레임씩 건너뛰며 비교해가는 방법이다. 실험에 의해 공간적 표본화를 통해서는 검출 시간이 크게 단축되지 않았다. 그러나 시간적 표본화를 통해 장면전환 검출 시간을 크게 단축시켜서 시간적 표본화 사용전보다 약 6 배의 검출 속도의 증가를 실현하였다.

동영상 검색 시스템에서 사용한 시간적 표본화 방법은 10 프레임씩 건너뛰면서 히스토그램 비교법을 적용한 것이다. 실험에 의해 시간적 표본화 간격을 10 프레임 정도로 하는 것이 검출 시간을 가장 적게 하는 것으로 나타났다. 비교하는 두 장면의 차이가 작으면 서로 같은 장면인 것으로 판단하고 계속 진행시켜가면서 비교한다. 만일 10 프레임 간격의 두 장면의 차이가 크면 그 내부에 장면 전환 지점이 있다는 가능성이 있는 것이므로, 이때에는 그 내부에 있는 장면들을 한 프레임씩 이동시켜가면서 비교함으로써 정확한 장면전환 지점을 찾는다.

시간적 표본화를 적용한 기존의 방법은 먼저 전체 영상을 10 프레임씩 건너뛰면서 히스토그램 비교를 하고 다시 영상의 처음으로 돌아가 장면 전환의 가능성이 있는 영역만 정밀 검사하는 두 패스에 의한 방법이다. 본 논문에서는 이 두 패스를 하나의 패스로 결합하여 정밀한 검사를 시간적 표본화와 동시에 수행하여 검출 속도를 좀더 향상시켰다. 또한 시간적 표본화를 하면서도 이중 비교법을 적용하여 점진적 변화 영상을 검출할 수 있도록 구현하였다.

III. 실험 결과 및 고찰

사용한 동영상의 화면 크기는 SIF(352*240) 포맷이며, 사용자가 선택한 MPEG 파일은 자동 인덱싱과정을 거쳐 표시창에 조각 그림으로 표시되며 검출하지 못한 장면이나 잘 못 검출된 장면은 수동으로 추가 또는 삭제할 수 있다. 인덱싱이 끝난 후에는 원하는 클립을 마우스로 클릭하면서 특정 클립 부분만 재생할 수 있다. 장면 검출이 끝나면 그 결과를 다음에 활용할 수 있도록 인덱스 파일에 저장할 수 있고 필요할 때 인덱스 파일을 불러다 쓸 수 있다.

자동 인덱싱 과정에 필요한 매개변수는 뉴스나 드라마 영상의 경우와 보안용 영상의 경우에 서로 다른 값을 설정하여 영상의 특성의 차이를 다소 극복하였다. 현재까지는 뉴스나 드라마 영상의 경우에는 검출 결과가 만족스러우나, 보안용 영상은 조명 및 다양한 장애 조건 때문에 검출 오류가 발생한다. 동영상 검색 프로그램에 영상의 장르를 선택할 수 있는 기능과 점진적 변화 영상 검출 여부를 선택할 수 있는 기능이 있다.

본 논문에서 구현한 동영상 검색 시스템을 개선할 점이 몇가지 있다. 현재 윈도 NT 가 설치된 PC 상에서 자동 인덱싱에 걸리는 시간은 실제 재생시간의 약 1.5 배가 소요되는데 이것을 단축시킬 필요가 있다. 이를 위해 MPEG 디코더 보드를 이용하면 현재 ActiveMovie 를 이용한 소프트웨어 디코딩을 사용하는 것으로부터 생기는 약점을 극복할 수 있을 것이다. 기존의 영상인식 방법에서는 카메라 플래쉬가 터지는 영상의 경우, 영상의 밝기가 급격히 변화하므로 검출 오류가 발생하는데, 본 논문에서는 인접 프레임간에 밝기가 연속적으로 급격히 변하는 경우에는 카메라 플래쉬가 터진 경우로 보아 하나의 장면으로 처리하여 인식률을 높였으며, 움직임이 많은 영상의 경우에는 카메라 움직임 분석 기법 및 물체의 움직임 추출 연구를 통해 검출 오류를 줄일 예정이다. 다음으로, 현재까지는 생성된 인덱스 파일 단위로만 검색이 가능하고 다양한 검색 기능이 부족하므로, 앞으로 데이터베이스와 검색 엔진의 구축이 필요하다.

VI. 결 론

본 논문에서는 MPEG 비트스트림을 분석하여 장면 전환 지점을 자동으로 찾는 실험을 워크스테이션을 통하여 시행하였으며 기존 실행한 실험을 바탕으로 PC 상에서 동영상 검색 시스템을 구현하였다

이를 통하여 방대한 양의 영상 자료를 디지털화하여 파일로 저장하고 영상에 관한 각종 정보를 데이터베이스로 구성한 뒤, 키워드 등을 사용하여 필요한 영상을 네트워크를 통하여 검색하고 이것을 편집 등에 활용할 수 있도록 하였으며, 영상의 자동 인덱싱 알고리듬을 개선하기 위해 다양한 연구가 필요하다.

참고문헌

[1] H. S. M. Beigi and S. H. Maes. Speaker, channel and environment[1] change detection. In *World Congress on Automation*, April 1998.

[2] G. J. Edwards, C. J. Taylor, and T. F. Cootes. Improving identification performance by integrating evidence from sequences. In *Proceedings of Computer Vision and Pattern Recognition*, pages 486487, 1999.

[3] K. Fukunaga. *Statistical Pattern Recognition*. AcademicPress, 2nd edition, 1990.

[4] C. Neti and A. W. Senior. Audio-visual speaker recognition for broadcast news. In *DARPA Hub 4 Workshop*, March 1999.