

논문-01-6-3-06

# 뉴스 정보의 단위 구조화를 위한 효율적인 앵커구간 추출 알고리즘

전승철\*, 박성한\*

## An Efficient Anchor Range Extracting Algorithm for The Unit Structuring of News Data

Seung-Chul Jun\* and Sung-Han Park\*

### 요 약

본 논문은 뉴스의 단위 구조화를 위해서 뉴스 동영상에 존재하는 앵커구간을 구분해내는 효율적인 알고리즘을 제안한다. 이를 위하여 본 논문은 단순히 장면 전환 지점을 이용하기 보다는 화면에 존재하는 앵커의 얼굴을 이용한다. 앵커구간에서는 앵커 얼굴 위치의 프레임 변화량을 계산하는 FRFD(Face Region Frame Difference)를 이용하여 앵커구간의 마지막을 찾는다. 한편 비앵커구간에서는 앵커 얼굴 추출을 이용하여 앵커구간의 시작점을 찾는다. 앵커 얼굴 추출은 처리 속도를 빠르게 하기 위해서 MPEG 동영상 부분 복호화를 통한 대략적 분석과 전체 복호화를 통해 얻어진 앵커 얼굴 후보를 검증하는 두 단계로 이루어져 있다. 이렇게 얻어진 앵커구간은 뉴스 분석의 기본 단계에서 이용이 가능하다. 특히 빠른 속도와 높은 recall 비율은 실제 뉴스 서비스 활용에 적당하다.

### Abstract

This paper proposes an efficient algorithm, extracting anchor ranges that exist in news video for the unit structuring of news. To this purpose, this paper uses anchors face in the frame rather than the cuts where the scene changes are occurred. In anchor range, we find the end position (frame) of anchor range with the FRFD(Face Region Frame Difference). On the other hand, in not-anchor range, we find the start position of anchor range by extracting anchors face. The process of extracting anchors face is consists of two parts to enhance the computation time for MPEG decoding. The first part is to find candidates of anchors face through rough analysis with partial decoding MPEG and the second part is to verify candidates of anchors face with fully decoding. It is possible to use the result of this process in basic step of news analysis. Especially, the fast processing and the high recall rate of this process are suitable to apply for the real news service.

## I. 서 론

최근 동영상 자동 분석에 관한 연구가 많이 진행되고 있다. 이는 컴퓨터 성능의 향상과 네트워크 속도의 향상으로 일반 사용자도 원격지에 있는 동영상을 볼 수 있는 환

경이 구축되었기 때문이다. 이러한 환경에서 많은 양의 동영상을 잘 정리하여 서비스 하기 위하여 여러 방향으로 연구가 되고 있다. 이러한 동영상 색인과 검색의 한 일환으로 MPEG-7에 대한 표준화가 진행 중 이고<sup>[1]</sup>, 또한 많은 회사에서 표준화에 앞서 실제 서비스 적용을 위한 여러 가지 방법을 제시하고 있다<sup>[2]-[5]</sup>. 이러한 동영상은 영상의 내용에 따라 다른 접근을 하게 된다. 특히 영화나 드라마, 스포츠 영상과 달리 뉴스는 영상 자체가 상당히 구조적으로 구성되어 있기 때문에 실제 서비스 시도를 가장

\* 한양대학교 공학대학 컴퓨터공학과

Dept. of Computer Science &amp; Engineering HanYang Univ.

※본 연구는 한국과학재단 목적기초연구(2000-2-303-005-3) 지원으로 수행 되었음

많이 하고 있다. 뉴스는 시간 순으로 앵커가 말하는 부분과 뒤이어서 나오는 비앵커 구간인 기자가 설명하는 화면으로 규칙적으로 구성되어 있다. 따라서 이를 하나의 뉴스 단위로 설정하여 1시간 분량의 뉴스를 여러 조각의 뉴스 영상으로 나누어 놓을 수 있다. 그림 1은 이러한 뉴스 동영상의 구조를 나타낸다. 이러한 뉴스 비디오의 구조를 잘 구성하기 위해서는 앵커 구간의 정확한 추출이 이루어져야 한다. 이러한 앵커구간의 추출을 위해서 여러 가지 방법이 제시되었다<sup>[6]</sup>.

동영상 분석을 위해서 가장 많이 이용되는 방법은 장면의 변화량을 이용하는 방법이다<sup>[7]</sup>. 이 방법은 현재 프레임과 이전 프레임의 화소값 차이와 히스토그램 차이 등과 같은 화면의 변화량을 하나의 수치로 표현할 수 있는 단위로 계산하여 그 변화량이 큰 지점을 장면의 바뀌는 지점으로 선택하는 방법이다. 이 방법을 이용하면 앵커구간에서 비앵커 구간으로 넘어가는 지점은 장면이 바뀌는 지점으로 볼 수 있기 때문에 간단히 앵커구간과 비앵커 구간을 구분할 수 있다. 하지만, 비앵커 구간에서 여러 번의 장면 변화가 일어나기 때문에 비앵커 구간에서 앵커구간으로 바뀌는 지점을 정확히 찾기는 쉽지 않다. 또한, 앵커구간 안에서 화면의 밝기가 심하게 변화하면 앵커구간에서 비앵커구간으로 진행된 것으로 잘 못 인식될 수 있고, 앵커구간 다음의 비앵커 구간의 밝기나 픽셀의 변화량이 심하지 않으면 비앵커 구간의 시작 지점을 놓칠 수도 있다.



그림 1. 뉴스 동영상의 일반적 구성  
Fig. 1. The general structure of news video data

이렇게 단순히 화면 변화량만을 가지고 앵커구간을 찾아내는 방법은 문제가 있기 때문에 앵커화면의 특성을 이용하는 방법이 연구되고 있다<sup>[6]</sup>. 그 중에서 앵커화면은 얼굴이 움직임이 없다는 점을 이용하여 앵커라고 추정되는 곳에 움직임 벡터를 얻어내어 움직임이 거의 없으면 앵커구간으로 설정하는 방법도 제안되었다<sup>[6],[8]</sup>. 하지만, 이러한 방법은 앵커 이외에도 움직임이 없는 영상이 존재할 수 있어 잘 못 인식할 가능성을 내포하고 있다. 또한 움직임 벡터를 얻는데 많은 계산량이 필요하기 때문에 성능을 저하시킬 수 있다. 더욱이 본 논문의 목적에 적용하기 힘든 부분은 앵커구간 외에 화면에 존재하는 자막정보, 앵커 얼굴 오른쪽에 존재하는 뉴스 아이콘 등의 추출에 더욱더

중심을 두고 있으므로 일반적으로 모든 뉴스에 대해서 앵커구간 추출이 불가능하다.

본 논문에서는 효율적으로 앵커 구간을 추출하기 위한 새로운 방법을 제안한다. 효율적 앵커 구간 추출을 위한 기본적인 방법은 앵커구간, 비앵커 구간의 프레임 차이를 따로 두는데 있다. 이 두 가지 프레임 차이는 앵커 구간과, 비앵커 구간에 적절한 별도의 척도를 가지고 실행된다. 이 방법은 뉴스 동영상의 구조적인 특징을 이용하고 있기 때문에 구조적 프레임 차이값이라 말한다. 또한 빠른 처리 속도를 위해서 처리 시간이 오래 걸리는 비앵커 구간 처리에서 계층적 검출방법을 사용하여 시간을 보다 절약한다.

이러한 방법을 설명하기 위해서 본 논문에서는 2장에서 구조적 프레임 차이값을 이용하여 앵커 장면을 검출하는 방법에 대해서 설명하고, 3장에서 앵커 얼굴을 찾아내는 방법을 알아본 다음 4장에서는 잘못된 앵커구간 제거를 위한 후처리를 수행하는 방법을 설명하며 5장에서 실제 뉴스 동영상 실험 결과를 보이고 마지막으로 6장에서 결론을 맺는다.

## II. 구조적 프레임 차이값을 이용한 앵커장면 검출

동영상 전 구간에 대해서 한가지 프레임 차이값을 이용하여 앵커구간을 정확히 찾아내기란 매우 어렵다. 왜냐하면 앵커구간의 동영상 특성은 다른 구간에 비해 한가지 특성으로 정의 내리기 쉽지만, 비앵커구간의 동영상 특성은 한 가지로 정의 내리기가 쉽지 않기 때문이다. 그러므로 본 논문에서는 앵커구간의 프레임 차이값 계산 방법과 비앵커구간 프레임 차이값 계산 방법을 달리하고 있다.

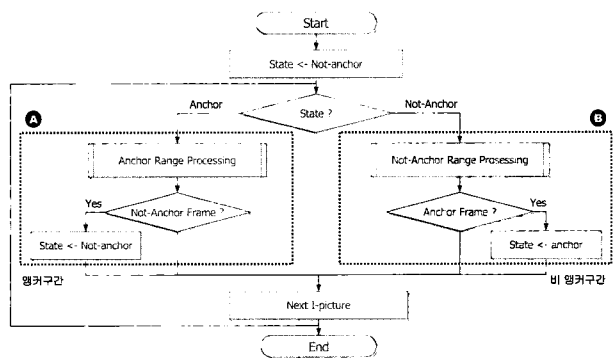


그림 2. 앵커구간 추출 알고리즘  
Fig. 2. Anchor range extraction algorithm

그림 2는 본 논문에서 사용하는 구조적 프레임 차이값 계산 방법이다. 앵커가 나타나는 장면을 앵커구간이라 하고, 앵커가 존재하지 않는 장면을 비 앵커 구간이라 한다. 이러한 구조는 앵커구간 동영상 특성과 비앵커 구간의 동영상 특성을 달리 설계할 수 있는 장점을 가져 다 준다. 앵커구간은 앵커 얼굴 영역을 기준으로 프레임 차이 값을 계산하고, 비 앵커 구간은 앵커 얼굴이 존재하는지 여부를 찾아내는 구간으로 정의한다. 뉴스 영상 분석의 시작은 항상 앵커구간이 아니라는 가정에서 시작한다. 앵커가 아닌 구간에서는 그림 2의 비앵커 구간을 처리한다. 비앵커 구간에서는 현재 프레임이 앵커구간의 시작인지 여부를 판단한다. 만약 앵커구간의 시작으로 판단되면 state 값은 Anchor로 설정되고 다음 단계에서 앵커구간 처리인 그림 2의 앵커 구간으로 이동한다. 또한 앵커 구간에서는 현재 프레임이 비 앵커 구간의 시작인가를 검사하여 계속적으로 앵커 구간을 처리를 수행할지 비앵커 구간 처리를 수행할지 여부를 결정한다.

1. 앵커 구간처리

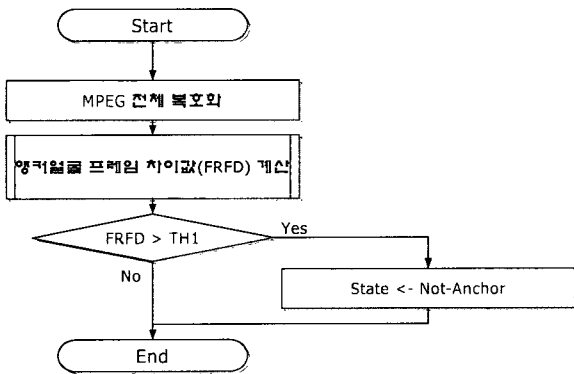


그림 3. 앵커 구간 처리 과정  
Fig. 3. The block diagram of anchor range process

앵커 구간 처리는 앵커 구간 화면에 존재하는 특징을 반영하여 정의한다. 앵커 구간 화면은 다음과 같은 4가지 특징을 가지고 있다.

- ① 앵커 얼굴이 반드시 나타난다.
- ② 앵커의 얼굴이 일정한 크기를 가지고 나타난다.
- ③ 얼굴의 움직임은 아주 미세하다.
- ④ 비디오 전반에 걸쳐 나타나는 배경이 유사하다.

이 중에서 앵커 구간 처리에서 사용하는 특징은 첫 번째, 두 번째, 세 번째 특징을 이용한다. 즉, 앵커얼굴이 반드시 나타나고, 얼굴의 크기가 변하지 않고, 그 얼굴의 움

직임이 거의 없는, 즉 변화량이 거의 없는 구간이 앵커 구간이라고 말할 수 있다. 따라서 본 논문에서는 앵커 구간에서는 화면 전체의 프레임 차이 값을 계산하기 보다는 앵커 얼굴이 존재하는 위치 부분만을 잘라내어 그 영역의 프레임 차이값을 얻어낸다. 이를 얼굴 영역 프레임 차이값 (Face Region Frame Difference, FRFD)이라 한다. 네 번째 특징은 오인식된 앵커구간을 제거할 때 사용한다.

그림 3은 FRFD를 계산하는 방법을 나타낸다. FRFD를 계산하기 위해서는 첫 번째 MPEG데이터 전체 복호화를 수행한다. 다음으로 전체 영상에서 얼굴 영역만을 오려낸다. 마지막으로 이전 프레임에서 얻은 얼굴 영상과 현재 얼굴 영상의 차이값을 얻는다. 그림 4는 FRFD를 얻을 때 사용하는 연속된 2개의 프레임을 나타낸다. 그림 4(a)는 뉴스 MPEG 동영상 중 앵커가 나타난 첫 번째 I-picture를 나타낸다. 그림 4(b)는 두 번째 I-picture를 나타낸다. 이 두 영상의 얼굴 영역은 거의 다른 것이 없기 때문에 얼굴이 존재하는 사각형 영역을 오려내어 두 영상의 차이를 계산하면 아주 적은 값을 갖게 된다.



그림 4. 앵커 영역 프레임 차이값 계산을 위한 얼굴 영역  
Fig. 4. The face region for calculating the frame difference in anchor block

영상의 차이값은 YCbCr컬러 모델에서 화소값 차이와 히스토그램 차이를 조합하여 [0, 1]사이로 정규화 한다. FRFD(얼굴 영역 프레임 차이값)는 다음 식 (1)에 의해서 주어진다.

$$FRFD = \alpha * PD_{fr}(t) + (1 - \alpha)HD_{fr}(t) \quad (1)$$

$$PD_{fr}(t) = \frac{1}{M^* N^* 255} \sum_{x=1}^M \sum_{y=1}^N FR(x, y, t-1) - FR(x, y, t) \quad (2)$$

$$HD_{fr}(t) = \frac{1}{M^* N} \sum_{i=1}^{255} H_{fr}(i, t-1) - H_{fr}(i, t) \quad (3)$$

$PD_{fr}$  : 얼굴 영역의 픽셀 차이값

$HD_{fr}$  : 얼굴 영역의 히스토그램 차이값

$FR$  : 얼굴 영역

$M$  : 얼굴 영역의 너비

$N$  : 얼굴 영역의 높이

$H$  : 얼굴 영역의 히스토그램

이 FRFD는 원칙적으로 프레임 차이값을 계산하는 방식을 그대로 따르고 있다<sup>7)11)</sup>. 다른 점은 프레임 전체에 적용하기 보다는 추출된 얼굴 영역에서만 그 계산을 수행한다는 점이 다르다. 식 (2)의 pixel difference는 Y, Cb 및 Cr영상 각각의 차이의 합을 정규화 한 값이다. 식 (3)의 histogram difference는 Y, Cb 및 Cr영상의 각각의 256개의 히스토그램의 차이를 합해서 정규화 한 값이다. FRFD는 식 (2)와 식 (3)을 가중치를 가지고, 더해서 정규화 한 값이다. 본 논문에서 사용한 가중치는 0.5를 사용한다. 이렇게 얻어진 FRFD 값은 앵커 구간에서는 거의 차이가 없기 때문에 0.05~0.2사이 값을 갖게 된다. 하지만, 비 앵커 구간의 화면의 FRFD는 0.3이상의 값을 갖기 때문에 쉽게 앵커 구간의 끝 지점을 찾을 수 있다. 그러나 이러한 FRFD값을 계산하기 위해서는 화면에 존재하는 얼굴 영역을 정확히 알고 있어야 한다. 얼굴 영역은 비 앵커구간 처리에서 앵커 구간의 시작을 찾아내면서 동시에 앵커 얼굴의 위치도 얻어낸다.

## 2. 비 앵커 구간 처리

비 앵커 구간 처리의 시작은 앵커 구간에서 현재 프레임이 더 이상 앵커구간이 아니라고 판단되면 시작된다. 비 앵커 구간 처리의 주 목적은 앵커 구간의 시작을 찾는 것이다. 앵커구간의 시작은 앞 장에서 설명한 앵커 화면의

특징을 이용한다. 그 중에서 앵커 화면에는 앵커 얼굴이 반드시 존재한다는 특징을 이용한다. 즉, 비 앵커 구간에서 앵커의 얼굴이 존재하면 그 프레임부터 앵커구간의 시작으로 가정한다. 앵커의 얼굴이 존재하는지 여부를 판단하기 위해서 정지영상에 존재하는 얼굴 추출 알고리즘을 이용하여 그 얼굴이 앵커인지를 판단한다. 최근 효율적인 얼굴 영역 추출 알고리즘이 많이 연구되어왔다[8]. 하지만, 이러한 일반적인 얼굴 영역 추출 알고리즘은 많은 계산량을 필요하기 때문에 모든 영상에 대해서 얼굴 영역 추출 알고리즘을 적용하여 동영상을 분석하는데 적합하지 않다. 따라서, 본 논문에서는 앵커 얼굴 제약 조건을 이용하여 얼굴 영역 추출의 정확도를 유지하면서 계산량은 줄이는 방법을 제안한다.

그림 5는 비앵커 구간의 처리과정을 나타낸다. 비 앵커 구간상에서 앵커 얼굴 추출 방법을 3 단계로 나누어서 수행한다. 이중 3 단계에서만 얼굴 영역 추출을 수행하도록 정의하고, 1, 2, 단계에서는 앵커 화면 가능성을 검사하도록 정의한다. 각 단계에서 앵커 화면일 가능성이 없으면 다음 프레임으로 진행한다. 각 단계는 다음과 같다.

### 1단계 : 이전 프레임과 현재 프레임의 프레임 차이값이 임계치 이상의 값인가를 검사

프레임 차이값이 작다는 의미는 이전 프레임과의 변화가 거의 없다는 의미이다. 즉, 이전 프레임이 앵커 프레임이 아니라면 현재 프레임도 앵커 프레임이 아니다. 이 단계는 복잡한 얼굴 영역 추출 알고리즘을 수행하기 전에 단순한 프레임 차이값 계산으로 다음 단계 수행의 여부를 결정한다. 여기서 사용하는 프레임 차이값은 계산량을 줄이기 위해 MPEG부분 복호화를 통한 Y, Cb, Cr의 macro block의 DC 계수값만을 얻은 데이터로 계산한다. 이 데이터는 원본 영상을 1/8로 축소한 영상이 되므로 부분 복호화를 통한 시간 이득과 작은 영상을 처리함으로써 얻어지는 시간 이득을 볼 수 있다.

### 2단계 : 얼굴 후보 영역을 추출

1단계에서 얻어진 1/8축소 영상에서 얼굴 후보영역을 추출한다. 얼굴 후보 영역을 추출하는 방법은 3장에서 설명한다. 후보영역 존재 여부에 따라 다음 단계의 진행 여부가 결정된다. 얼굴 후보 영역이 존재하지 않으면 3단계를 수행하지 않고 다음 프레임으로 진행한다. 만약 얼굴 후보 영역이 한 개 이상이면 3단계를 수행한다.

**3단계 : 실제 얼굴 영역을 추출**

MPEG전체 복호화를 통한 완전한 정지 영상을 얻은 후 후보 영역에 대응하는 영역들을 오려낸다. 오려진 각 영역별로 얼굴이 존재하는지를 다시 한번 검사한다. 이 과정에서 1/8축소 영상에서 얻어진 후보 영역 중 잘못된 얼굴 영역들이 제거되고, 순수한 앵커 얼굴만이 남게 된다. 이 과정에서 후보 영역이 모두 앵커 얼굴이 아닌 경우 다음 프레임으로 진행한다. 만약 후보 영역이 앵커 얼굴로 판정되면 현재 프레임을 새로운 앵커 구간의 시작으로 설정한다. 또한 여기서 얻은 얼굴 위치를 이용하여 다음 프레임의 FRFD를 계산한다.

이와 같은 과정은 모든 프레임에 얼굴 영역 추출을 수행하기 보다는 얼굴이 존재할 가능성이 많은 프레임에 대해서만 얼굴 영역을 추출하기 때문에 많은 계산의 이득을 얻을 수 있다.

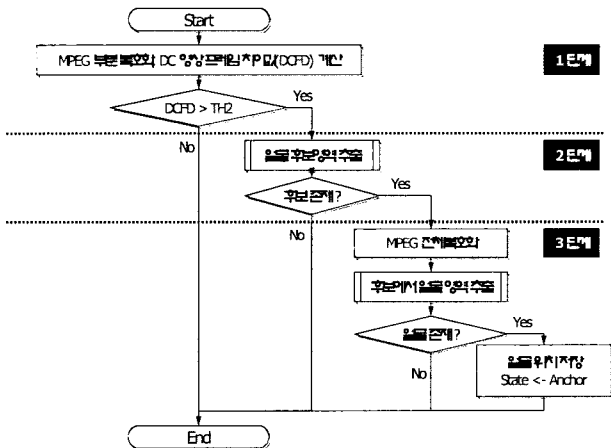


그림 5. 비앵커구간 처리 과정  
Fig. 5. The block diagram of not-anchor range process

그림 5의 1단계 조건에서 TH2의 값은 [0, 1]사이의 값을 갖는다. 이 TH2값은 전체 프로세스의 성능에 크게 영향 미치는 값이 된다. 이 값이 클 경우 후보가 많아지고 잘못된 후보가 많아진다. 값이 작을수록 후보가 적어지고 놓친 후보가 많아진다. 비디오 동영상 분석의 특성으로 볼 때 처리 속도 보다는 정확도가 중요하다. 정확도 중에서도 recall rate값이 높을수록 실제 활용에 적합하다. 즉, 잘 못 찾아진 것은 사람이 쉽게 제거 가능하지만, 놓친 것이 사람이 다시 찾아내기가 어렵다. 결과적으로 TH2값을 적게 하는 것이 전체적인 속도를 저하 시키지만, 실제 활용에 적합 하다고 볼 수 있다. 하지만 너무 작게 하면 후보 데

이터가 많아서 전체적인 처리 속도가 떨어진다. 본 논문에서는 실제 뉴스 데이터를 근거로 실험적으로 얻어진 데이터에 의해서 TH2값을 0.3으로 설정한다. 이 때 놓친 후보는 적으면서 전체 후보의 개수가 가장 적게 나타난다. 그림 5의 2 단계 조건에서 후보의 존재는 3장의 얼굴 영역 추출결과 얼굴 존재 여부로 판단한다. 얼굴이 존재하면 후보로 간주하고, 얼굴이 존재하지 않으면 다음 단계를 진행한다. 후보 얼굴을 얻는 과정은 앵커 얼굴이 아닌 객체를 제거하는데 있다. 만약 앵커 얼굴이 아니면서 앵커 얼굴 후보를 나타내는 경우발생하면 다음 과정에서 제거된다. 후보 객체를 얻는 과정은 3장의 얼굴 색상을 기준으로 이진 영상을 얻은 후 얻어진 이진 영상 중에서 다음 조건을 가지는 객체를 제거하여 앵커 얼굴 후보를 얻는다.

- ① 객체의 면적이 15픽셀 이하인 객체
- ② 작은 영상의 상하좌우 테두리에 붙어 있는 객체
- ③ 작은 영상을 정확히 4등분하여 오른쪽 상단 영역에 포함되는 객체

①번 조건은 앵커 얼굴 객체의 면적은 적어도 15픽셀보다 크므로 이 보다 작은 객체는 제거하는 조건이다. ②, ③번 조건은 얼굴 객체가 앵커 얼굴이라는 조건이므로 테두리에 붙어 있는 객체는 앵커얼굴이 아니다. 또한 오른쪽 상단 영역에 포함되는 객체는 앵커 얼굴 보다는 앵커 오른쪽에 나타나는 뉴스 정보 영상이므로 이러한 위치에 있는 객체는 제거하는 조건이다.

**III. 앵커 얼굴 영역 추출**

얼굴 영역 추출 방법은 얼굴을 이용한 영상처리를 위한 기본적인 알고리즘으로 많은 연구가 있었다<sup>[9]</sup>. 또한, MPEG-4의 객체 단위 코딩을 위한 연구도 많이 이루어지고 있다. 하지만, 이러한 방법은 임의의 영상에서 얼굴을 찾아내기 위한 방법이기 때문에 정확도를 높이기 위해서는 많은 계산량이 필요한 연산을 수행해야 한다. 하지만 본 논문에서 사용하는 방법은 동영상 의미 분석을 처리하는 단계의 일부분이기 때문에 최대한 빠르게 분석한 후 서비스를 제공하여야만 된다. 따라서, 본 논문은 일반적인 얼굴이 아닌, 앵커 얼굴인 점을 이용하여 앵커 얼굴 제약 조건을 정의하여 앵커 얼굴을 얻어낸다. 앵커 얼굴의 추출은 앵커 구간의 시작을 알아내는 중요한 처리이기 때문에 매우 정확하게 얼굴 영역을 추출해야 한다. 또한 동영상에서 추출해야 하기 때문에 가능한 적은 계산량으로 영역을 추출해야 한다. 본 논문은 이러한 조건을 만족시키는 앵커

얼굴 영역 추출 방법을 제안한다.

최근 얼굴 추출 알고리즘은 얼굴 색상을 이용하여 좋은 결과를 얻고 있다<sup>[8]</sup>. 본 논문에서도 기본적으로 얼굴 색상을 이용한 얼굴 추출 알고리즘을 사용한다. 그리고, 마지막에서 앵커 얼굴 제약 조건을 이용하여 앵커 얼굴을 얻어낸다. 얼굴 영역 추출 알고리즘은 그림 6과 같다.

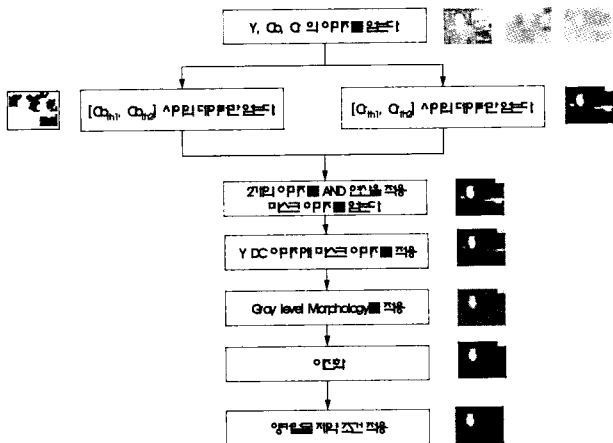


그림 6. 얼굴 영역 추출 알고리즘  
Fig. 6. The face region extraction algorithm

얼굴 색상은 YCbCr컬러 모델의 Cb, Cr부분에서 일정한 범위에 존재하는 특징을 가진다<sup>[10]</sup>. 이를 이용하여Cb와 Cr 픽셀 데이터의 특정 범위만을 얻어낸다. 첫 번째로 얼굴일 가능성이 있는 영상을 얻기 위해서 Cb는 [77, 127]사이의 데이터와 Cr은 [133, 173]사이의 데이터를 얻는다. 두 번째로 얻어진 2개의 데이터에서 동시에 존재하는 영역만을 추출 하기위해 2개의 이진 영상을 AND연산을 한 후보 얼굴 영역을 얻어낸다. 세 번째 단계에서 얻어진 데이터를 Y영상에 마스크를 씌워서 얼굴 후보 영상을 얻어낸다. 네 번째에서 영상에 존재하는 얼굴 보다 작은 잡음들을 제거 하기 위해서 그레이레벨 모폴로지 Opening연산을 수행한다. 마지막으로 이진화를 거쳐 앵커얼굴 제약 조건을 적용하여 얼굴 영역을 추출한다. 본 논문에서 사용한 앵커 얼굴 제약 조건은 다음과 같다.

- ①  $width - width * \acute{a} < \text{앵커 얼굴을 둘러싸는 사각형의 넓이} < width + width * \acute{a}$   
width : 객체를 둘러싸는 사각형의 너비  
 $\acute{a}$  : 얼굴 넓이 변위 비율
- ②  $height - height * \acute{a} < \text{앵커 얼굴을 둘러싸는 사각형의 넓이} < height + height * \acute{a}$   
height : 객체를 둘러싸는 사각형의 높이

- $\acute{a}$  : 얼굴 높이 변위 비율
- ③ 얼굴 객체의 면적  $> S$   
S : 얼굴 최소 면적
- ④ 좌우가 대칭이다.
- ⑤ 얼굴 크기가 일정
- ⑥ 얼굴 위치가 거의 변화가 없다.
- ⑦ 얼굴 주변의 영상은 거의 변화가 없다.

앵커얼굴 추출 과정에서 사용되는 앵커 얼굴 제약 조건은 (1)-(5)번까지 이고, (6), (7)번 제약조건은 앵커 구간 추출에서 사용한다. 얼굴의 좌우 대칭값의 정보는 얼굴 후보 영역의 그레이 영상에서 얻어낸다. 이때 하나의 대칭값을 얻는 것 보다는 영상의 중심을 기준으로 좌우 여러 개의 대칭값을 얻어 최대 값을 얻는다. 이렇게 함으로써 얻어진 얼굴 영상이 오른쪽이나 왼쪽으로 조금 치우친 영상에서도 높은 대칭 값을 얻게 되어 얼굴 후보 영역의 미세한 오차를 보완할 수 있다.



그림 7. 얼굴 후보 영역 추출  
Fig. 7. The extraction of face candidate regions

이러한 얼굴 영역 추출 방법은 앞 비앵커구간 처리 방법의 2단계, 3단계에서 사용된다. 즉, 비 앵커 구간에서 부분 복호화를 통하여 축소된 작은 영상을 처리한 다음 후보 영역이 존재할 경우 전체 복호화를 통하여 전체 영상을 얻어서 다시 한번 수행한다. 이것은 처음부터 전체 영상에 대하여 수행하기 보다는 축소 영상에서 추출된 후보 영역을 잘라내어 잘라진 영역에 대해서만 처리를 수행하여 처리 시간의 이득을 보고, 후보 영역에서만 얼굴을 찾아 좀더 정확한 얼굴을 추출하기 위함이다. 그림 7은 얼굴 영역 추출 과정을 보여준다. 이때 축소영상은 원래 영상의 1/8로 축소한 영상이기 때문에 후보 영역을 8배 확대한 위치가 앵커 얼굴의 후보 영역으로 볼 수 있다. 다음으로 후보영역의 미세한 오차를 줄이기 위해서 120% 증가된 사각형영역을 최종 후보 영역으로 설정한다. 그림 7(c)는 최종 후보영역에서 다시 얼굴을 추출하는 과정을 보이고 있다.

#### IV. 잘못된 앵커구간 제거

본 논문에서 사용하는 앵커구간 추출 알고리즘은 앵커 얼굴이 존재하는 구간을 앵커구간으로 설정하기 때문에 앵커 얼굴과 유사하게 나타나는 비앵커 구간에서 앵커로 오 인식하는 경우가 있다. 따라서 이러한 오 인식된 앵커 구간을 반드시 제거해 주어야 한다. 오 인식된 앵커구간을 제거하기 위해서 본 논문은 세 가지 특징을 이용한다.

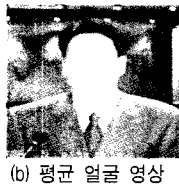
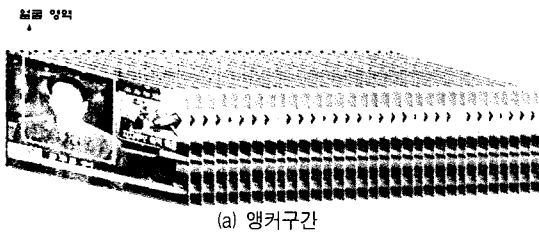


그림 8. 앵커 구간 대표 영상 추출 과정  
Fig. 8. The representative images of anchor range

첫 번째는 앵커구간의 길이를 이용한다. 길이가 임계치 이하인 구간은 오 인식된 구간으로 판정한다. 임계치는 2 초 이하의 구간으로 사용한다. 이는 모든 앵커 구간이 2 초 이상 나타난다는 특징을 이용한 것이다. 두 번째는 앵커구간의 얼굴 차이값 들의 분포를 이용한다. 하나의 앵커구간에서 앵커 얼굴 차이값의 평균값은 많아야 0.2이하의 값을 갖는다. 즉, 얼굴 차이값 평균이 0.2이상인 구간은 오 인식된 구간으로 판정한다. 세 번째는 앵커 구간들의 대표 영상들의 차이값을 이용한다. 이는 다른 앵커 구간이라도 동일한 앵커가 나타날때 앵커 얼굴 대표 영상들의 차이값은 작게 되는 특징을 이용한 것이다. 앵커구간 대표 영상은 앵커구간의 얼굴을 평균하여 얻는다. 그림 8은 앵커구간을 대표하는 이미지를 얻는 과정을 나타낸다. 그림 8(a)에 나타난 앵커 구간에서 앵커 얼굴 주변의 평균 영상을 얻은 결과가 그림 8(b)이다.

이렇게 얻어진 앵커구간의 대표 영상들을 이용하여 각 대표 영상의 차이를 이용하여 비슷한 영상끼리 모으는 작

업을 수행한다. 수행결과 오직 하나의 영상이 존재한 그룹은 오 인식된 앵커 구간으로 판정한다. 그림 9은 이 과정을 나타낸 흐름도이다.

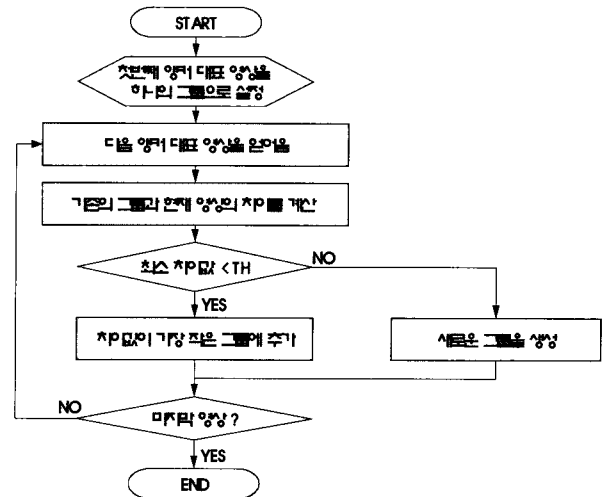


그림 9. 앵커 구간 대표 영상을 이용하여 오 인식 구간 제거 과정  
Fig. 9. The elimination procedure of false detection using representative image of anchor range

#### V. 실험 결과

본 논문에서 사용한 실험 데이터는 3개 방송국(KBS, MBC, SBS)의 메인 뉴스를 사용한다. 각 방송국 별로 2일 간의 뉴스 영상을 얻어서 사용한다. 동영상은 MPEG-1으로 부호화되고 352\*240크기의 해상도를 갖는 영상을 사용한다. GOP구간의 크기는 6혹은 15프레임으로 구성된다. GOP구조는 IBBPBBP로 진행되는 영상을 사용한다. 본 논문에서 사용한 뉴스 MPEG 데이터는 <http://vipmc.hanyang.ac.kr/newsdata/>에서 얻을 수 있다. 실험 환경은 Pentium III 700Mhz이다.

그림 10은 여러 방송국 뉴스의 각각 다른 4명의 앵커의 얼굴을 추출한 결과를 보이고 있다. 각 앵커 별로 앵커얼굴 후보 영상부터 마지막 실제 추출 결과까지 7개의 영상이 연산의 순서대로 나열되어 있다. 첫번째 후보 영상은 2.2장에서 얻어진 후보 객체에 해당하는 영역만을 얻어낸 영상이다. 중간 단계의 영상 중 두 번째, 세 번째 영상은 Cb[77, 127], Cr[133, 173]사이의 이진영상을 말하고, 네 번째 영상은 두 번째, 세 번째 영상을 AND연산을 수행한 후 그 결과를 후보에 마스크를 씌운 결과이다. 다섯 번째, 여섯 번째는 그레이 모폴로지를 적용한 후 이진화를 한 결과

이고, 마지막은 최종적으로 얼굴 영역만을 추출한 결과이다. 실험에서 보는 바와 같이 서로 다른 동영상의 다른 얼굴을 가지는 앵커의 얼굴이 정확히 추출됨을 볼 수 있다.

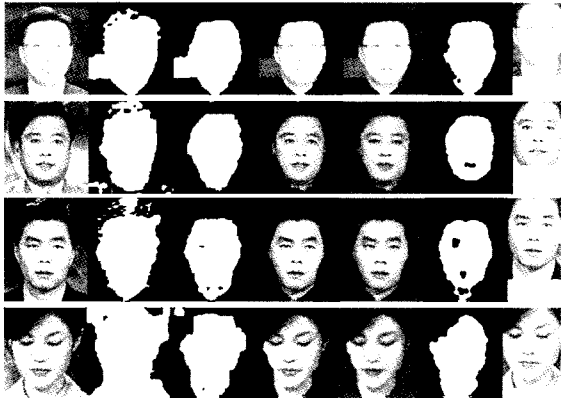


그림 10. 앵커 얼굴 추출 결과  
Fig. 10. The results of extracting anchor face

그림 11은 앵커 구간과 비 앵커 구간의 프레임 차이 값을 달리 계산한 결과를 보여주고 있다. 앵커 구간이 끝나는 지점에 앵커 얼굴 차이값이 크게 커짐을 알 수 있다. 또한 비 앵커구간에서 앵커 얼굴 후보가 나타났지만, 최종적으로 앵커 얼굴 특징을 이용하여 앵커 화면이 아님으로 판단한 경우를 보인다. 이렇게 얻은 앵커 구간 중에는 앵커 얼굴과 비슷한 구간이 앵커 구간으로 오 인식된 결과가 포함되어 있다. 이렇게 오 인식된 구간을 제거한 결과의 일부가 그림 12이다.

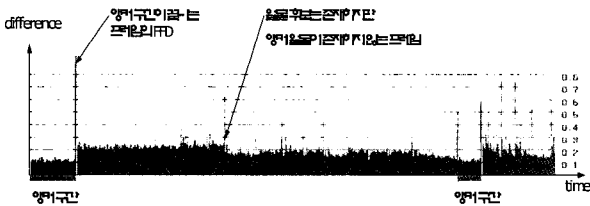


그림 11. 앵커 구간과 비 앵커구간의 프레임 차이 값 계산  
Fig. 11. The frame differences of anchor range and not-anchor range

그림 12에서 각 행의 첫번째 영상은 그룹을 대표하는 영상이 되고, 나머지는 앵커 구간을 대표하는 영상이다. 영상 아래 있는 숫자는 그룹을 대표하는 영상들과의 차이값 중 최소 값을 표현한 값이다. 처음 2개의 그룹의 영상들은 앵커이기 때문에 여러 개의 영상이 그룹을 이루고 있다. 나머지 그룹 오 인식된 영상이기 때문에 오직 한 개의 영상으로 그룹이 이루어진다. 하나의 영상으로 이루어진 그

룹을 모두 제거하여 오 인식된 앵커 구간을 제거한다.

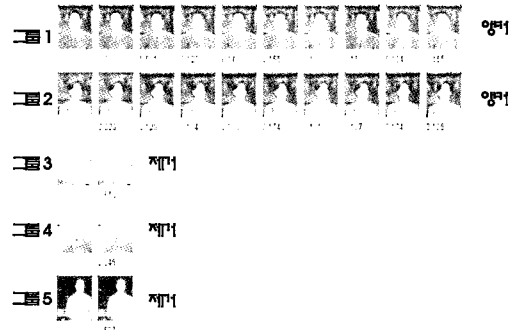


그림 12. 잘 못 인식된 앵커구간 제거 결과  
Fig. 12. The result of eliminating false anchor range

표 1. 앵커 구간 추출 결과  
Table 1. The result of extracting anchor block

방송국	날짜	길이	분석 시간	A:	B:	C:	오인식된 앵커 구간	놓친 앵커 구간	Precision (C/B)	Recall (C/A)
				실제 앵커 구간	추출된 앵커 구간	적절한 앵커 구간				
KBS	2001. 2.24.	41분 39초	4분 6초	26	26	0	0	0	100%	100%
	2001. 4. 9.	56분	8분 10초	35	38	34	4	1	89.5%	97.1%
MBC	2000. 3.12.	41분 19초	4분 22초	25	26	25	1	0	96.2%	100%
	2000. 3.15.	49분 22초	5분 19초	33	38	32	6	1	84.1%	97.0%
SBS	2001. 3.15.	34분 25초	5분 41초	20	21	19	2	1	90.5%	95%
	2001. 3.24.	31분 20초	5분 10초	16	16	16	0	0	100%	100%
평균									93.4%	98.2%

표 1은 일반 방송국의 메인 뉴스를 실험한 결과를 나타낸 것이다. Precision은 추출된 앵커 구간과 적절한 앵커 구간의 비율을 나타낸다. 그리고 Recall은 전체 적절한 앵커 구간과 검색된 적절한 앵커 구간의 비율을 나타낸다. 앵커 구간 추출은 Precision보다 Recall비율이 높은 것이 더욱 더 실제에 적용에 적합하다. 이는 앵커 구간 보다 더 많이 추출된 결과에서 잘못 된 구간을 제거하는 것은 쉬우나, 누락된 앵커 구간을 사람이 찾는 것은 많은 시간이 걸리기 때문이다. 본 논문의 결과는 Precision비율은 93.4%인 것에 비해 Recall비율은 98.2%로 높은 것으로 볼 수 있다. Precision비율이 낮은 이유는 가능한 앵커 구간과 비슷한 구간은 모두 추출하기 때문에 정밀도가 떨어지고 있다. 하지만, 실제 앵커 구간보다 더 많은 구간을 앵커구간으로 추출하여 놓친 앵커구간은 거의 존재하지 않음을



볼 수 있다. 오 인식된 경우는 화면의 사람 얼굴이 앵커 얼굴 제약 조건을 만족하는 경우에 오 인식 될 가능성을 내포하고 4장에서 제거되지 못 할 경우 앵커구간으로 오 인식된다. 이렇게 오 인식된 앵커구간은 발생빈도는 실험 결과에 의해 7.9%정도 발생한다. 하지만 발생된 오 인식 구간은 실제 활용에서 사람에 의해 쉽게 제거가 가능하기 때문에 실제 활용에 적합하다.

## VI. 결론

본 논문에서는 뉴스 동영상을 의미적인 뉴스 clip의 구조화를 위한 앵커구간과 비앵커구간을 추출한다. 이는 단순히 앵커만을 추출하는 경우와는 비앵커구간의 시작 정지 영상을 이용하여 전체 뉴스를 대표하는 영상을 제시할 수 있다. 기존의 장면 전환 지점을 기준으로 앵커 구간을 추출하는 방법에 비해 본 논문에서 제안하는 방법은 앵커 얼굴 이 존재하는 구간을 앵커구간으로 정의 함으로써 앵커구간 추출의 정확도를 높인다. 또한, 영상 전체의 장면 변화량을 계산하기 보다는 얼굴 부분의 픽셀의 변화량을 계산하여 앵커 구간을 추출한다. 본 논문에서는 빠르고 효율적인 분석을 하기 위하여 MPEG영상의 전체 복호화가 아닌 부분 복호화를 통한 단계적 분석법을 제시하여 전체 분석 속도를 향상 시킨다. 실험 결과 각 동영상 재생 시간에 비해서 1/10시간 안에 뉴스 동상에서 앵커구간을 추출한다. 더욱이 디지털 방송이 현실화 되는 시점에서 MPEG동영상 분석은 더욱 더 필요로 하다. 본 논문에서는 동영상 표준인 MPEG동영상을 실험 데이터로 사용함에 따라 실제에 적용하기에 매우 적합하다.

## 참 고 문 헌

- [1] Thomas Sikora, "MPEG Digital Video-Coding Standards," *IEEE Signal Processing*, pp.82-100, Sep. 1997.
- [2] The Informedia Project information : Digital Video Library Proposal URL (<http://www.informedia.cs.cmu.edu/html/description.html>)
- [3] H.D. Wactlar, T.Kanade, M.A. Smith, and S.M. Stevens, "Intelligent Access to Digital Video : Informedia Project," *IEEE Computer Magazine*, pp. 46-52, May 1996.
- [4] VideoLogger, Media Analysis Software, Virage, <http://www.virage.com/>
- [5] Savitha Srinivasan, Dulce Ponceleon, Arnon Amir, and Dragutin Petkovic, "What is in that video anyway?: In Search of Better Browsing," *Multimedia*, vol.I, pp. 388-393, 1999.
- [6] 이종구, 양명섭, 유철중, 방옥배, "자동인택싱에 기반한 뉴스 비디오 검색 시스템의 설계 및 구현", *정보과학회 논문지*, C호, pp.43-52, 1999년 2월
- [7] Seungchul Jun, Sunghan Park, "An Automatic Cut Detection Algorithm Using Median Filter And Neural Network," *ITC-CSCC 2000*, vol.II, pp.1049-1052, Jul. 2000.
- [8] D. Chai and K. N. Ngan, "Face Segmentation Using Skin-Color Map in Videophone Applications," *IEEE Transaction on Circuits and Systems for Video Technology*, vol.9, no.4, pp.551-564, Jun. 1999.
- [9] Roberto Brunelli and Tomaso Poggio, "Face Recognition: Feature versus Templates," *IEEE Transacion on Pattern Analysis and Machine Intelligence*, vol.15, no.10, pp.1042-1052, Oct. 1993.
- [10] Jing Huang, S Ravi Kumar, Mandar Mitra, Wei-Jing Zhu, and Ramin Zabih., "Image Indexing Using Color Correlogram", *International Conference on Computer Vision and Pattern Recognition, IEEE*, 1997.
- [11] B. L. Yeo and B. Liu, "Rapid scene change detection on compressed video," *IEEE Transaction on Circuit and Systems for Video Technology*, vol.5, no.6, pp.533-544, Dec. 1995.
- [12] Y. Nakajima, "Universal scene change detection on MPEG-coded data domain," *Visual Communications and Image Processing '99*, 12-14 San Jose, California, vol.3024, Feb. 1997.

저 자 소 개



전 승 철

1996년 2월 : 한양대학교 전자계산학과 졸업(공학사)  
 1998년 2월 : 한양대학교 대학원 전자계산학과 졸업(공학석사)  
 1999년 2월~현재 : 한양대학교 대학원 전자계산학과 박사과정  
 주관심분야 : Video Indexing, MPEG-4, MPEG-7, Computer Vision, Human Face Recognition



박 성 한

1970년 2월 : 한양대학교 전자공학과 졸업(공학사)  
 1973년 8월 : 서울대학교 대학원 전자공학과 졸업(M.S.)  
 1984년 5월 : 텍사스주립대학 전기 및 컴퓨터공학과 졸업(Ph. D.)  
 1974년 3월~1978년 8월 : 경북대학교 전자공학과 전임강사  
 1984년 5월~1984년 8월 : 미국 텍사스 주립대학 Instructor  
 1984년 8월~1986년 2월 : 금성사 중앙 연구소 수석 연구원  
 1989년 8월~1990년 7월 : 미국 텍사스 주립대학 Visiting Researcher  
 1995년 3월~1997년 2월 : 한양대학교 공학대학 학장  
 1986년 3월~현재 : 한양대학교 전자계산학과 교수  
 주관심분야 : 멀티미디어 통신, 컴퓨터 비전, B-ISDN