

# MPEG 압축 비디오 상에서의 얼굴 영역 추출 및 인식

## Face Detection and Recognition in MPEG Compressed Video

여 장 옥\*      유 명 현\*\*  
(Chang-Yu Lu) (Myung-Hyun Yoo)

**요 약** 본 논문에서는 MPEG 압축 비디오 상에서 얼굴 영역을 추출하고 이를 인식하는 방법에 대하여 제안한다 제안된 방법은 크게 MPEG 압축 비디오의 처리를 위한 축소된 DC 영상의 구성 단계, 축소된 DC 영상에서의 얼굴 영역 추출 단계, 그리고 얼굴 영역이 추출된 프레임에 대한 압축 복원 및 얼굴 인식의 3단계로 구성되어있다

DC 영상의 구성 단계에서는 압축 복원 없이 DCT 계수의 DC 값과 2개의 AC 값만을 사용하여 부분적인 2차원 역 DCT 변환을 이용한 방법을 사용하였으며, 얼굴 영역 추출 단계에서는 DC 영상에 대해 얼굴의 색상 및 형태 정보를 이용한 얼굴 후보 영역 추출 방법과 K-L 변환 및 역 변환의 오차에 의한 얼굴 영역 추출 방법을 사용하였다 얼굴 인식 단계에서는 얼굴 영역이 추출된 프레임에 대하여 GOP 단위의 압축 복원을 수행한 후 고유 얼굴 영상을 이용한 방법으로 얼굴 인식을 수행하였다

제안된 방법의 성능을 검증하기 위하여 뉴스와 드라마 MPEG 비디오를 대상으로 실험을 수행하였으며, 실험 결과 제안된 방법이 효율적임을 알 수 있었다

**주제어** 얼굴 인식, 얼굴 영역 추출 MPEG 압축 비디오

**Abstract** In this paper we present a face recognition and face detection algorithm in MPEG compressed video. The proposed method consists three stage of processing steps. The first step is to produce a spatially reduced DC image form MPEG compressed video for processing And the second step is face detection on reduced DC image Finally, the last step is face recognition on partially extracted compressed frames which contain the detected faces

The spatially reduced DC image is produced from two dimensional inverse DCT of the DC coefficient and the first two AC coefficients The face detection is performed on DC image and face recognition is performed on one extracted frame per GOP by using the K-L transform

In order to evaluate the proposed method, we carried out experiments on video database The experiment results show the proposed method is very efficient and helpful for target tasks

**Keywords** Face Recognition, Face Detection, MPEG compressed video

### 1. 서론

최근 대용량 멀티미디어 데이터베이스의 증가와 통신 및 디지털 미디어 처리 기술의 발달로 인하여 비디오 자동 색인 및 검색 기술에 대한 요구가 증가하고 있다 현재까지의 연구는 주로 비디오 데이터에 대해 사람이 직접 색인을 수행하고, 사용자가 주제어에 기반하여 검색하는 텍스트 기반의 검색 방법에 대

고려대학교 인공시각연구센터  
연구세부 분야 : 얼굴 인식, 동영상 처리  
주제어 : 얼굴 인식, 얼굴 영역 추출 MPEG 압축 비디오  
우편 번호 : 136-701  
주소 : 서울시 성북구 안암동 고려대학교 인공시각연구센터  
전화 : 3290-4283  
FAX : 926-2168

\* cylvu@image.korea.ac.kr

\*\* mhvoo@image.korea.ac.kr

해 이루어졌으나, 비디오 데이터 내의 각 객체들을 자동으로 분리하고, 이들의 특징을 추출하여 색인 및 검색에 사용하는 내용기반 검색 방법에 대한 연구가 선진국을 중심으로 활발하게 진행되고 있다. 그러나 비디오 내에 존재하는 모든 객체를 분리하는 데에는 한계가 있으며, 이를 모두 색인화하는 것 또한 거의 불가능하다. 따라서 비디오 내에 존재하는 여러 객체들 중에서 가장 중요한 객체를 분리하고 특징을 추출하여 색인화하는 접근 방법이 제안되고 있다. 사람의 얼굴은 비디오 내에 존재하는 가장 중요한 객체 중의 하나이며, 이를 분리하여 색인으로 사용할 경우 얼굴 영상을 사용한 질의 및 검색이 가능하고, 얼굴 영상을 중심으로한 비디오 데이터의 구조화가 가능하다.

또한 방대한 양의 비디오 데이터를 저장하고 처리하기 위해서는 비디오의 압축이 필요하고, 이의 관한 연구 또한 필수적이다. 전 세계적으로 여러가지 압축 방법이 제안되고 있으나, 국제표준기구의 MPEG이 동영상 압축의 국제 표준으로 받아들여지고 있다. 이미 많은 양의 비디오 데이터가 MPEG 표준 형식으로 저장 되어있으며 앞으로도 거의 모든 비디오 데이터가 MPEG 형식을 사용할 것으로 예상된다. 이러한 압축된 비디오 데이터를 처리하기 위하여 현재까지는 모든 압축을 해제한 후 처리를 수행하고, 이를 다시 압축하는 방법을 사용하였으나, 처리 시간의 문제와 저장 공간의 낭비로 인하여 압축된 상태에서의 정보만을 사용하여 비디오 데이터를 처리할 수 있는 기술도 절실히 요구되고 있다.

따라서 본 논문에서는 MPEG 압축 비디오 상에서 효과적으로 얼굴 영역을 추출하고, 추출된 얼굴 영역의 프레임만을 압축 복원하여 얼굴 인식을 수행하는 방법에 대해 제안한다. 제안된 방법은 크게 압축 비디오로부터의 축소된 DC 영상의 구성 단계, 축소된 DC 영상에서의 얼굴 영역 추출 단계 그리고 얼굴 영역이 추출된 프레임을 복원하여 인식을 수행하는 얼굴 인식의 3단계로 구성된다. DC 영상의 구성 단계에서는 DCT 계수의 DC 값과 2개의 AC 값만을 사용하여 부분적인 2차원 역 DCT 변환을 이용한 방법을 사용하였으며, 얼굴 영역 추출 단계에서는 앞에서 얻어진 DC 영상에 대해 얼굴의 색상 및 형태 정보를 이용한 얼굴 후보 영역 추출 방법과 K-L 변환 및 역변환의 오차에 의한 정확한 얼굴 영역 추출 방법을 사용하였다. 얼굴 인식 단계에서는 얼굴 영역이 추출된 프레임에 대하여 GOP 단위의 압축 복원을 수행한 후 고유 얼굴 영상을 이용한 방법으로 얼굴 인식을

수행하였다. 제안된 방법의 순서도는 그림 1과 같다.

본 논문의 구성은 다음과 같다. 2장에서는 압축 비디오 상에서의 영상 처리와 얼굴 영역 추출 및 인식에 관련된 기존의 연구들에 관하여 살펴보고, 3장에서는 DC 영상의 구성 방법에 관해 그리고 4장에서는 얼굴 영역 추출 및 인식 방법에 관해 설명한다. 5장에서는 제안된 방법의 성능을 검증하기 위한 실험 및 그 결과를 분석하고, 마지막으로 6장에서는 결론 및 향후 연구 방향에 대하여 논의한다

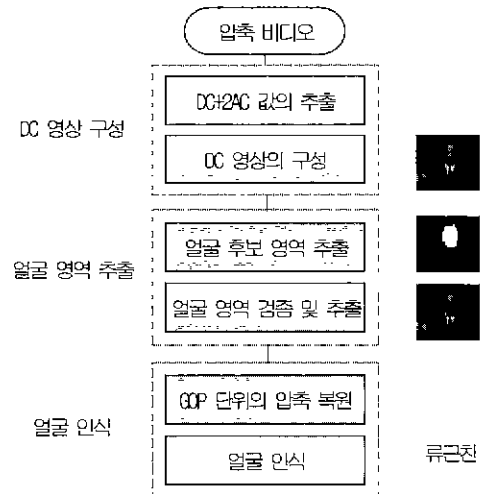


그림 1 제안된 방법의 순서도

## 2. 관련 연구

### 1.1. 압축 비디오 처리

일반적인 압축 비디오의 처리 과정은 복호화를 통하여 압축을 해제한 후 이를 처리하고 다시 압축하는 방법이 사용되어왔다. 그러나 복호화 과정으로 인한 처리 시간과 저장 공간의 낭비로 인하여 현재는 주로 압축된 상태에서 직접 처리하는 방법에 관한 연구가 진행되고 있다. 현재 가장 많이 사용되는 동영상 압축 방법으로는 MPEG, JPEG-비디오, H 261 등이 있으며, 특히 비디오 CD 중심의 저장 미디어 동영상 압축 표준인 MPEG을 중심으로한 압축 비디오에서의 처리 기법의 관한 연구가 진행되고 있다. MPEG 압축 비디오에서 사용할 수 있는 특징들로는 DCT 계수 또는 움직임 벡터 등이 있으며, 이들 특징을 이용한 영상 처리 및 비디오 합성, 비디오 색인 및 검색, 섯

경계 검출, 자막 추출, 얼굴 영역 추출에 관한 연구가 진행되고 있다

영상처리 및 비디오 합성에 관한 연구로는 B Shen[1] 등 및 S. F. Chang[2] 등을 중심으로한 DCT 영역에서의 처리 기법 등이 소개되고 있다 B Shen 등의 경우 DCT Convolution을 정의하고 이를 응용한 여러 가지 비디오 합성 및 특수 효과들을 압축된 상태에서 수행하는 방법에 대하여 제안하였다 S F Chang 등의 경우 압축 비디오 상에서의 다양한 영상 편집 및 합성에 관한 방법에 대해 제안하였다 MPEG 으로 압축된 비디오에 대하여 새로운 압축 해제 방법 및 MC(Motion Compensated) DCT 비디오 구성 방법을 제안하였으며, 이를 사용하여 다자간 화상 통신에 사용될 영상 합성에 적용하였다. 또한 블록 단위의 DCT 계수 연산을 통하여 디졸브(Dissolve), 페이드 인(Fade In), 크로마키(Chroma Key) 등의 특수 효과 및 비디오 버퍼 제어 방법을 통한 GOP 단위의 비디오 편집 방법을 제안하였다.

비디오 색인 및 검색에 관한 연구로는 S. F. Chang 등의 내용기반 비디오 검색 시스템인 VideoQ에 관한 연구가 소개되고 있다[3]. VideoQ는 압축 비디오 상에서 자동으로 의미 있는 객체를 분리하고 이를 색인화 하여 저장 및 검색에 활용하는 시스템이다. 비디오 내의 객체에 대한 색상, 질감, 모양 및 움직임(방향 및 지속 시간)을 색인의 특징으로 삼고 사용자는 이들 특징들의 적절한 조합을 스케치하여 질의를 수행할 수 있다.

셋 경계 검출에 관한 연구로는 DCT 계수를 이용한 Meng 등의 방법[4]과 Kobola 등의 방법[5]이 있으며 Yeo의 DC 영상 구성에 의한 방법[6]이었다 또한 Wei와 Zhong 등은 MPEG 비디오에서 움직임 벡터를 사용하여 셋의 경계 프레임에서 B-프레임의 움직임 벡터가 대부분 0 값을 갖는다는 사실에 기반하여 양의 값을 갖는 움직임 벡터의 수가 임계값 이하인 프레임을 셋의 경계 프레임으로 추출하는 방법을 제안하였다[7]. 자막 추출에 관한 연구로는 DC 영상 내에서 화소간 명도차를 이용한 Yeo의 방법[6]이 연구되고 있으며, 얼굴 영역 추출에는 Wang의 DCT 계수 및 색상 정보에 의한 방법[8]이 연구되고 있다

## 1.2. 얼굴 영역 추출 및 인식

얼굴 영역 추출은 정확한 얼굴 인식을 위한 핵심 기술로써, 얼굴 인식의 연구화 더불어 활발한 연구가 진행되어 왔다. 그러나 복잡한 배경에서의 얼굴 영역

추출 기술은 90년대 들어서야 본격적으로 연구되기 시작하였으며, 실시간 얼굴 영역 추출에 관한 연구는 90년대 중반에 들어서야 비로소 진행되기 시작했다. 현재까지 연구된 대표적인 얼굴 영역 추출 방법으로는 형판정합에 의한 방법[9], 퍼지 이론에 의한 방법[10], 신경망을 이용한 방법[11], 얼굴의 색상 및 형태 정보를 이용한 방법[12], K-L 변환에 의한 방법[13] 등이 있다 또한 Dai와 Nakano가 제안한 특정한 제한 조건에서의 얼굴 영역을 결정하는 방법[14] 등이 소개되고 있으며, 그 외에도 다양한 방법들이 연구되고 있다.

형판정합에 의한 방법과 퍼지이론에 의한 방법은 단순한 배경에 대해 뛰어난 성능을 나타내나 복잡한 배경을 갖는 영상이나 어린이, 대머리 얼굴과 같이 형판과 잘 일치하지 않는 영상의 경우 성능이 떨어지는 단점을 가지고 있다 신경망을 이용하는 방법의 경우 얼굴의 기술어집이나 표정 변화 등에 대처할 수 있는 유연성과 빠른 추출 속도를 내는 장점이 있지만, 얼굴 영역이 아닌 일반 배경을 학습시킬 적절한 예가 없다는 문제점이 있다. 얼굴의 색상 정보를 이용한 방법은 구현이 용이하고 얼굴의 위치 및 크기를 한번에 결정할 수 있는 장점이 있으나, 다양한 피부색 변화와 조명의 변화에 의한 오류가 많다는 단점을 가지고 있다.

K-L 변환을 사용한 방법은 비교적 안정적인 성능을 나타내며 대머리 영상 등 특수한 영상에 대해서도 얼굴 영역의 추출이 가능하나 많은 처리 시간이 필요하다는 문제점이 있다

얼굴 인식은 얼굴의 표정, 조명, 시점의 변화에 무관하면서도 각 개인을 구별할 수 있는 특징 값을 얻어내어 얼굴을 분류하는 기술이다. 현재까지의 얼굴 인식에 관한 연구는 주로 소규모 데이터베이스를 대상으로 하고 있으며, 실 시간처리 문제에 많은 관심을 기울이고 있다. 대표적인 얼굴 인식 방법으로는 Konen과 Ekkehard[15]의 Gabor wavelet을 사용한 방법, Turk와 Pentland[13]의 K-L 변환을 사용한 방법, 신경망을 이용한 방법[16] 그리고 특징 기반의 인식 방법 등이 있다.

Konen과 Ekkehard의 Gabor wavelet을 사용한 방법의 경우 얼굴의 크기, 각도, 표정 등에 무관한 인식이 가능하나 한 영상을 인식하기 위한 처리 시간이 많이 소요되어 대용량의 데이터베이스를 대상으로 하는 얼굴 영상 검색에는 부적합한 방법이다 Turk와 Pentland가 제안한 K-L 변환을 사용한 방법의 경우

얼굴 표정, 각도, 조명 등에 둔감하지만 얼굴 크기 변화에는 대단히 민감한 인식 방법으로 대용량의 얼굴 영상 데이터베이스에 적용하여 실험한 결과, 우수한 성능을 나타내었으나 실험에 사용한 데이터가 단일 배경, 단일 조명의 영상인 경우였다. 신경망을 이용한 방법의 경우 소규모 얼굴 영상에 대해서는 좋은 성능을 나타내나, 인식 대상 얼굴의 수가 많아짐에 따라 인식 성능이 급격히 저하되는 문제점을 갖고 있다. 특징 기반의 인식 방법은 다양한 얼굴 특징을 이용한 복합적인 인식 방법으로 뛰어난 인식률을 나타내지만 각 개인을 구별하기 위한 파라미터의 종류가 많고 유사도 측정 과정이 복잡하여 대용량 얼굴 영상 검색에는 비효율적인 방법이다.

### 3. DC 영상의 구성

MPEG 압축 비디오에서 사용할 수 있는 특징 값들은 DCT 계수와 움직임 벡터 등이 있다. DCT 계수는 화면 내의 화소 정보를 주파수 영역에서 나타내는 것으로 I 프레임에서 얻어질 수 있으며, 움직임 벡터는 화면간의 유사도 및 움직임을 나타내는 척도로서 P 또는 B 프레임에서 얻어질 수 있다.

본 논문에서는 압축된 비디오 데이터를 직접 처리하기 위하여 DCT 계수를 사용한 DC 영상의 구성 방법을 사용한다. DC 영상은 압축된 비디오의 모든 정보가 아닌 DCT 계수의 일부분만을 사용하여 구성된 영상으로 원 영상 보다 낮은 해상도의 축소된 영상이다. DC 영상은 DCT 계수의 DC 값 또는 DC+2AC 값을 사용하여 구성되며 DC 값만을 사용할 경우 원 영상의 1/64, DC+2AC 값을 사용할 경우 1/16의 크기를 갖게된다. 본 논문에서는 DC+2AC 값을 사용하여 DC 영상을 구성하였다. 또한 I 프레임 이외에 P 프레임 및 B 프레임에서의 DC 영상 구성을 위하여 움직임 벡터 정보 및 예측 프레임 내에서의 DCT 계수 정보를 사용하여 근사값을 얻어내는 방법을 사용하였다. DC 영상의 구성 방법을 사용할 경우 MPEG 비디오 데이터를 압축 복원하지 않고도 영상내의 중요한 정보를 얻어낼 수 있으므로 매우 효율적이다.

DC 영상의 구성은 DCT 계수의 DC+2AC 값을 사용한다. MPEG 압축의 특징인 프레임 유형에 따른 압축 방법의 차이로 인하여 DC 영상의 구성 방법 역시 다음과 같이 프레임 유형 별로 다르게 수행된다.

#### 3.1 I 프레임의 경우

I 프레임의 경우 DCT 계수를 직접 추출할 수 있으

므로 각 블록에 대하여 DC 및 2개의 AC 성분을 추출하여 다음과 같은 2차원 역 DCT 변환을 통해 4x4 크기의 DC 영상 블록을 얻어낼 수 있다.

$$\frac{1}{8} \begin{bmatrix} d+a_1+a_2 & d-a_1+a_2 \\ d+a_1-a_2 & d-a_1-a_2 \end{bmatrix}$$

여기서  $d, a_1, a_2$  는 각각 DCT 계수의 DC 성분과 2개의 AC 성분을 나타낸다.

#### 3.2 P 또는 B 프레임의 경우

P 와 B 프레임의 경우 움직임 벡터를 사용한 예측 부호화를 사용하므로 DCT 계수를 직접 구할 수 없다. 따라서 이 경우에는 그림 2와 같이 움직임 벡터 정보와 예측 프레임 내의 DCT 계수를 사용한 근사값을 사용하게 된다.

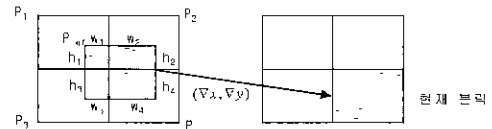


그림 2 예측 프레임 및 움직임 벡터 정보

즉 다음과 같이 예측 블록내의 DCT 계수는 주위 4 블록과의 겹쳐지는 비율에 따라 결정됨을 의미한다.

$$(DCT(P_{ref}))_{00} \approx \frac{1}{64} \sum_{i=1}^4 h_i w_i (DCT(P_i))_{00}$$

여기서  $(DCT(P_i))_{00}$  는  $i$  번째 블록에서의 DCT 계수를 의미한다.

### 4. 얼굴 영역 추출 및 인식

얼굴 영역의 추출은 앞에서 얻어진 DC 영상에서 이루어진다. 본 논문에서는 빠르고 정확한 얼굴 영역의 추출을 위하여 두 단계의 얼굴 영역 추출 방법을 제안한다. 우선 첫 번째 단계에서는 색상 및 형태 정보에 기반하여 얼굴 후보 영역을 설정하였다. 두 번째 단계에서는 K-L 변환 및 역변환의 오차를 이용한 얼굴 영역 추출 방법으로 최종적인 얼굴 영역을 추출하였다. 이처럼 두 단계의 처리 방법을 사용함으로써 처리 시간을 단축시킬 수 있었으며, 정확도를 향상시킬 수 있었다.

### 4.1 얼굴 후보 영역의 추출

얼굴 영상이 배경과 확실히 구분되는 특징으로는 피부색 및 눈, 코, 입 등의 얼굴 구성요소와 이들의 상대적인 위치 관계를 들 수 있다 이 중에서도 특히 피부색은 다른 특징들에 비하여 배경과 확실히 구별되는 특징으로서, 얼굴 후보 영역 추출을 위하여 널리 사용되고 있다.

본 논문에서는 얼굴 후보 영역의 추출을 위하여 얼굴의 색상 및 형태 정보를 사용한다. 색상 정보를 이용한 방법은 MPEG의 휘도 신호인 Y 성분 이외에 색차 신호인  $C_b$ ,  $C_r$  성분에 기반한 이진화 방법을 사용하였다 즉 피부색에 해당하는 색차 신호의 임계값을 적용하여 얼굴 후보 영역을 설정하는 것이다 MPEG의 4:1:1 표준 형식의 경우 4개의 휘도 신호에 대하여 하나의 색차 신호가 할당되므로 설정된 후보 영역은 DC 영상에서 4개의 화소가 기본 단위가 된다.

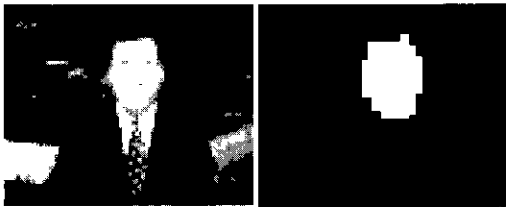


그림 3 얼굴 후보 영역 추출의 예

설정된 후보 영역에 대해서 얼굴의 형태 정보에 기반한 얼굴의 크기, 가로세로비의 임계치를 적용하여 이를 크게 넘어서거나 못 미치는 경우를 제외하여 최종적인 얼굴 후보 영역을 설정하였다

### 4.2 얼굴 영역의 검증 및 추출

얼굴 영역의 검증 및 추출은 각 후보 영역에 대하여 K-L 변환 및 역변환의 오차값을 이용한 방법을 사용하였다. 즉 그림 4와 같이 임의의 영상을 고유 얼굴 영상들로 이루어진 축으로 사영한 후 이를 다시 역변환 했을 경우, 고유 얼굴 영상들과의 사영으로 인하여 얼굴 영상과 기타 영상들이 각각 다른 오차값을 갖는 성질을 이용한 방법이다

얼굴 영역은 각 후보 영역에 대하여 오차값을 구하고 최소값을 갖는 영역으로 설정한다

정확한 얼굴 영역의 추출은 각 후보 영역에 대하여 다해상도 탐색을 수행하므로써 얼굴의 크기 및 위치를 결정하는 방법을 사용하였다

### 4.3 얼굴 인식

얼굴 영상을 인식하기 위한 방법으로는 고유 얼굴 영상을 이용한 방법을 사용하였다 얼굴 영역 추출 단계에서 얼굴 영역의 크기와 위치를 결정한 후 얼굴의 크기를 일정하게 정규화하여 K-L 변환을 수행하였다 변환의 결과로서 변환의 결과로서 주성분 값이 추출되고 이 주성분 값을 바탕으로 얼굴 영상을 인식하였다

얼굴 인식을 수행하기 위하여 DC 영상으로부터 얻

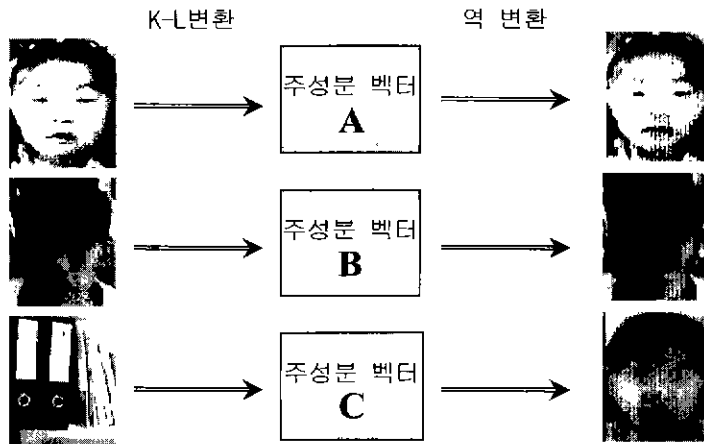


그림 4 K-L 변환의 오차값을 이용한 얼굴 영역 추출 방법

굴 영역이 추출된 프레임에 대하여 GOP 단위의 압축 복원을 수행한다 또한 연속적으로 추출된 얼굴 영역에 대해서는 프레임 유형에 따른 우선 순위와 GOP 단위의 시간 관계를 고려하여 얼굴 인식을 수행하였다. 즉 I, P, B 3가지의 프레임 유형에 대하여  $D P > B$ 의 우선 순위를 설정하고, 얼굴 영상이 1초내에 다른 인물로 변화하지 않는 가정하에 하나의 GOP 내에서 우선 순위가 높은 하나의 프레임만을 압축 복원하여 얼굴 인식에 사용하였다.

또한 일반적으로 대부분의 얼굴 영상이 I 프레임에서 검출되므로 GOP 단위의 압축 복원시 첫 번째 I 프레임만을 복원하는 것으로 얼굴 인식이 가능하다. 따라서 얼굴을 인식하기 위하여 모든 프레임을 압축 복원하지 않고 GOP 당 하나의 프레임만을 압축 복원하므로 압축 비디오에서의 효율을 향상시킬 수 있었다.

얼굴 인식은 고유 얼굴 영상을 이용한 방법을 사용하였다. 고유 얼굴 영상을 이용한 방법이란 얼굴 영상의 K-L 변환으로 얻어진 주성분 값을 이용한 얼굴 인식 방법이다. 정규화된 얼굴 영상에 K-L 변환을 수행하여 주성분 값을 얻어내고 이를 얼굴 영상들의 특징 값인 주성분 값들과의 MSE(Mean Square Error)를 기준으로 영상들을 인식하였다

이 방법의 특징으로는 얼굴의 표정 및 조명 변화에 무관한 인식이 가능하다는 점과 빠른 인식 속도 등이

며, 각 개인에 대해 몇 개의 주성분 값만을 인식 파라미터로 사용하므로 대용량 얼굴 인식 및 검색에 가장 적합한 방법이라 할 수 있다 그러나 얼굴 크기에 민감한 변화를 보이는 단점을 지니고 있어 정확한 얼굴 영역의 추출이 요구된다

본 논문에서는 얼굴 인식을 수행하기 위하여 200장의 표본 영상을 선정하고, K-L 변환을 수행하여 상위 40개의 주성분 값을 인식에 사용하였다

## 5. 실험 및 결과 분석

### 5.1 실험 환경

제안된 방법을 검증하기 위한 실험에는 Pentium 166MHz의 PC를 사용하였으며, 352x240의 해상도를 갖는 MPEG-1 형식의 압축 비디오를 사용하였다. 실험에 사용된 비디오 데이터는 TV 뉴스와 드라마로서 Optibase MPEG Fusion System MPEG-2 Encoder를 사용하여 채록되었으며, 제안된 방법의 성능을 검증할 수 있도록 다양한 크기, 표정, 시점을 갖는 얼굴 영상을 포함하고 있다. 그림 5는 실험에 사용된 비디오 데이터의 예이다.

### 5.2 실험 및 결과 분석

표 1과 표 2는 제안된 방법에 대한 실험 결과이다



그림 5 실험에 사용된 비디오 데이터의 예

표 1은 얼굴 영역 추출에 대한 실험 결과로서 여기서 N은 얼굴 영상이 포함된 GOP의 수를 나타내며 NC는 정확하게 얼굴 영역이 추출된 GOP의 수를 나타낸다 그리고 NFN은 얼굴 영역을 검출해내지 못한 GOP의 수를 NFP는 잘못 검출된 얼굴 영역의 GOP 수를 의미한다 표 2에서는 얼굴 영역이 추출된 GOP에 대하여 얼굴 인식을 수행한 결과로서 N은 얼굴 영역이 추출된 GOP의 수를 NC는 정확하게 얼굴이 인식된 GOP의 수를 나타내며 NF는 얼굴 인식이 잘못된 GOP의 수를 나타낸다

종류	N	NC	NFN	NFP
뉴스 1	137	102	35	40
뉴스 2	120	94	26	22
드라마	90	71	19	34

표 1 얼굴 영역 추출 실험 결과

얼굴 영역 추출 결과의 경우 단순한 배경에서 일정한 크기를 갖고 정면을 바라보는 얼굴 영상에 대해서는 좋은 성능을 보였으나 복잡한 배경하에서 얼굴의 시점이 심하게 변화된 영상을 검출해내지 못하였다 또한 축소된 DC 영상을 사용하여 얼굴 영역을 결정하므로 크기가 작은 얼굴 영상의 경우에는 얼굴 위치에 약간의 오차가 발생하였다.

종류	N	NC	NF
뉴스 1	102	88	14
뉴스 2	94	70	24
드라마	71	65	6

표 2 얼굴 인식 실험 결과

얼굴 인식 결과의 경우 대부분의 얼굴 영역이 I 프레임에서 연속적으로 검출되어 동일 인물에 대한 반

복적인 얼굴 인식을 수행하는 경우가 많았다. 얼굴 인식의 성능은 대부분 얼굴 영역 추출 성능에 의해 좌우되었으며, 압축 복원된 상태에서 얼굴 인식을 수행하므로 얼굴의 크기가 작거나 각도가 심하게 변하지 않았을 경우에는 좋은 인식 성능을 나타내었다.

## 6. 결론 및 향후 연구 계획

본 논문에서는 MPEG 압축 비디오 내에서 가장 중요한 객체인 얼굴 영상을 추출해내고 이를 인식하는 방법에 대하여 제안하였다. 제안된 방법은 크게 MPEG 압축 비디오의 처리를 위한 DC 영상 구성 단계, DC 영상에서의 얼굴 영역 추출 단계, 그리고 얼굴 영역이 추출된 비디오에 대한 압축 복원 및 인식의 3단계로 구성되어있다

MPEG 압축 비디오에서 압축 복원 없이 정보를 얻어내기 위하여 DC 영상 구성 방법을 사용하였다. MPEG 압축의 특징인 프레임 유형에 따른 압축 방법의 변화를 고려하여 I 프레임과 P, B 프레임에서 각각 다른 DC 영상 구성 방법을 사용하였다.

I 프레임에서의 DC 영상 구성은 각 블록 단위로 DCT 계수중 DC+2AC 값만을 추출하여 2차원 역 DCT 변환 방법을 사용하여 구성하였으며, 프레임간 예측을 사용하는 P 및 B 프레임에 대해서는 각 블록에 대하여 움직임 벡터와 참조되는 블록의 DCT 계수를 사용한 근사값을 사용하였다.

얼굴 영역의 추출에서는 얼굴 영상의 두드러진 특징 중 하나인 얼굴의 색상 및 형태 정보를 사용하여 얼굴 후보 영역을 설정하고, 설정된 후보 영역에 대하여 K-L 변환의 오차값을 사용하여 얼굴 영역을 추출하는 두 단계의 얼굴 영역 추출 방법을 사용하였다.

얼굴 인식 단계에서는 DC 영상으로부터 얻어진 얼굴 영역이 추출된 프레임에 대하여 GOP 단위의 압축 복원을 수행하고, 이를 통해 얻어진 영상에 대하여 얼굴 인식을 수행하였다. 여기서 프레임 유형에 따른 우선 순위를 설정하고 얼굴 영상이 1초 이내에 변하지 않는다는 가정하에 하나의 GOP에 대해 한번의 얼굴 인식을 수행하였다. 얼굴 인식에 사용된 얼굴의 특징으로는 조명 및 표정 변화에 무관한 K-L 변환의 주성분 값을 사용하였다. 얼굴 영상을 인식하기 위하여 200장의 표본 영상을 구성하고 K-L 변환을 수행하여 상위 40개의 주성분 값을 특징 값으로 사용하였다. 얼굴 영상의 유사도 측정은 MSE를 기준으로 하였다.

대용량 멀티미디어 서비스에 대한 요구가 급격히

증가하고 있는 현재의 추세에 따라 비디오 데이터를 효율적으로 관리하고 처리할 수 있는 내용 기반 비디오 검색에 대한 연구의 필요성이 날로 증가하고 있다. 압축 비디오 내에서의 얼굴 영역 추출 및 인식에 대한 연구는 이러한 내용 기반 비디오 검색의 큰 영향을 미칠 것이다. 또한 대부분 압축되어 저장되는 비디오를 보다 효율적으로 처리할 수 있도록 압축 비디오에 대한 특징 추출 및 색인 방법의 연구가 진행되어야 할 것이며, 보다 일반화된 얼굴 영역 추출 및 인식을 위해서는 다양한 비디오 편집 효과와 여러 가지 카메라 동작에 의한 변화를 수용할 수 있도록 비디오 분석 기술의 연구가 함께 진행되어야 할 것이다.

### 감사의 말씀

본 연구는 정보통신부 대학기초연구지원사업의 연구비 지원을 받았음

### 참고 문헌

- [1] B. Shen, I. K. Sethi and V. Bhaskaran, "Digital Video Blue Screen Editing in Compressed Domain," Proc. of SPIE Conf on Visual Communication and Image Processing'97, San Jose, CA, Vol. 3024, pp. 971-982, 1997.
- [2] S. F. Chang and D. G. Messerschmitt, "A New Approach to Decoding and Compositing Motion-Compensated DCT-Based Images," Proc of ICASSP'93, Minneapolis, MN, April, 1993
- [3] S. F. Chang, W. Chen, H. J. Meng, H. Sundaram, and D. Zhong, "VideoQ - An Automatic Content-Based Video Search System using Visual Cues," Proc. of ACM Multimedia Conf Seattle, WA, Nov, 1997
- [4] J. Meng, Y. Juan and S. F. Chang, "Scene Change Detection in a MPEG Compressed Video Sequences," Proc of SPIE Conf. on Digital Video Compression : Algorithms and Technologies, Vol 2419, San Jose, pp. 14-25, 1995
- [5] V. Kobola, D. S. Doermann and K. L. Lin, "Archiving, Indexing, and Retrieval of Video in the Compressed Domain," Proc. of SPIE Conf on Multimedia Storage and Archiving Systems, Vol 2916, San Jose, pp. 78-89, 1996.
- [6] B. L. Yeo, "Visual Content Highlighting via Automatic Extraction of Embedded Captions on MPEG Compressed Video," Proc. of SPIE Conf. on Digital Video Compression: Algorithms and Technologies, San Jose, CA, Vol 2688, pp. 38-47, 1996
- [7] Q. Wei and Y. Z. Zhong, "Content-Based Parsing in Video Database," Proc. of First Int Conf on Multimodal Systems and Technologies, 1997.
- [8] H. Wang and S. F. Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video," IEEE Trans. on Circuit and Systems for Video Technology, special issue on Multimedia Systems and Technologies, 1997
- [9] 이 미숙, 이 성환, "얼굴 영상의 이해를 위한 복잡한 배경에서의 정면 얼굴 영역의 실시간 추출," 한국정보과학회 가을학술발표회 발표논문집, pp 609-612, 1996년 10월
- [10] Q. Chen, H. Wu and M. Yachida, "Face Detection by Fuzzy Pattern Matching," Proc. of 5th Int Conf. on Computer Vision, pp. 591-596, 1996
- [11] H. A. Rowley, S. Baluja and T. Kanade, "Neural Network-Based Face Detection," Proc. of IEEE Computer Society Conf on Computer Vision and Pattern Recognition, pp 203-208, 1996.
- [12] K. Sobottka and I. Pitas, "Extraction of Facial Regions and Features using Color and Shape Information," Proc. of Int. Conf on Image Processing, pp. 483-486, 1996.
- [13] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, Vol 3, No. 1, pp. 71-86, 1991
- [14] Y. Dai and Y. Nakano, "Extraction of Facial Images from Complex Background



- using Color Information and SGLD Matrices," Proc of Int Workshop on Automatic Face-and Gesture-Recognition, pp. 238-242, 1995.
- [15] W. Konen, and S. Ekkehard, "ZN-Face: A System for Access Control using Automatic Face Recognition," Proc of Int Workshop on Automatic Face-and Gesture-Recognition, pp 18-23, 1995.
- [16] S. H. Lin, S. Y. Kung and L. J. Lin, "Face Recognition/Detection by Probabilistic Decision-Based Neural Network," IEEE Trans on Neural Networks, Vol 8, No 1, pp. 114-132, January, 1997