

# MPEG 압축 비디오 상에서의 얼굴 영역 추출 및 인식

여창욱<sup>†</sup>, 황본우<sup>‡</sup>, 이성환<sup>†‡</sup>

<sup>†</sup>고려대학교 영상정보처리학과, <sup>‡</sup>고려대학교 컴퓨터학과  
E-mail: {cvlu, bwhwang, swlee}@image.korea.ac.kr

## Face Detection and Recognition in MPEG Compressed Video

Chang-Yu Lu<sup>†</sup>, Bon-Woo Hwang<sup>‡</sup>, Seong-Whan Lee<sup>†‡</sup>

<sup>†</sup>Dept. of Visual Information Processing, Korea University

<sup>‡</sup>Dept. of Computer Science and Engineering, Korea University

### 요약

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 구성되어 있다. 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.

### 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 대해 이루어졌으나 비디오 데이터 내의 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 분리하는데에는 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 중심으로 한 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 대

이터가 MPEG 표준 형식으로 저장 되어있으며 앞으로는 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 처리할수있는 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 역 변환의 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.

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 방법에 관해 설명한다 5장에서는 얼굴 인식 방법에 대해 설명하고 마지막으로 6장에서는 제안된 방법의 성능을 검증하기 위한 실험 및 그 결과에 대하여 논의한다

## 2. 관련연구

### 2.1 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 압축 비디오에서 사용할 수 있는 특징들로는 DCT 계수 또는 움직임 벡터 등이 있으며, 이들 특징을 이용한 셋 경계 검출, 자막 추출 얼굴 영역 추출에 관한 연구가 진행되고 있다 셋 경계 검출에 관한 연구로는 DCT 계수를 이용한 Meng 등의 방법[1]과 Kobola 등의 방법[2]이 있으며 Yeo의 DC 영상 구성에 의한 방법[3]이 있다 자막 추출에 관한 연구로는 Yeo의 방법[3], 그리고 얼굴 영역 추출에는 Wang의 DCT 계수 및 색상 정보에 의한 방법[4]이 연구되고 있다.

### 2.2 얼굴 영역 추출 및 인식

얼굴 영역 추출은 배경과 얼굴 영역을 분리하는 기술로써, 정확한 얼굴 인식을 위한 핵심 기술이다 복잡한 배경에서의 얼굴 영역 추출 기술은 90년대 들어서야 본격적으로 연구가 시작 하였으며 실시간 얼굴 영역 추출에 관한 연구는 90년대 중반에 들어서야 비로소 진행되기 시작했다. 대표적인 얼굴 영역 추출 방법으로는 평균정합에 의한 방법[5], 신경망을 이용한 방법[6], 얼굴의 색상 정보를 이용한 방법, K-L 변환에 의한 방법[7] 등이 있으며, 이외에도 다양한 방법들이 연구되고 있다

얼굴 인식은 얼굴의 표정, 조명 시점의 변화에 무관하면서도 각 개인을 구별할 수 있는 특징 값을 얻어내어 얼굴을 분류하는 기술이다 대표적인 얼굴 인식 방법으로는 Konen과 Ekkelhard[8]의 Gabor wavelet을 사용한 방법, Turk와 Pentland[7]가 제안한 K-L 변환을 사용한 방법, 특징 기반의 인식 방법 등이 있다

## 3. DC 영상의 구성

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 갖게 된다. 본 논문에서는 DC+2AC 값을 사용하여 DC 영상을 구성하였다 DCT 계수를 사용한 부분적인 2차원 역 DCT 변환은 다음과 같다

$$\frac{1}{8} \begin{pmatrix} d+a_1+a_2 & d-a_1+a_2 \\ d+a_1-a_2 & d-a_1-a_2 \end{pmatrix} \quad (1)$$

여기서  $d, a_1, a_2$ 는 DCT계수의 DC 성분과 2개의 AC 성분을 나타낸다 I 프레임의 경우 역 DCT 변환을 위한 계수를 쉽게 구할 수 있으나, P와 B 프레임의 경우 움직임

벡터를 사용한 예측부호화를 사용하므로 DCT 계수를 직접 구할 수 없다. 따라서 이 경우에는 움직임 벡터 정보와 예측 프레임 내의 DCT 계수를 사용한 근사값을 사용하게 된다 [3]

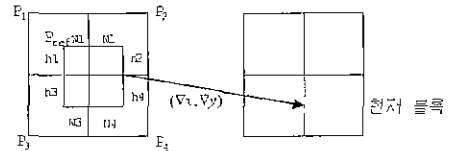


그림 1. 움직임 벡터 및 예측 프레임의 DCT 계수

$$(DCT(P_{r,c}))_{00} \approx \frac{1}{64} \sum_{i=1}^4 h_i u_i (DCT(P_i))_{00} \quad (2)$$

이거시  $DCT(P_i)$ 는 1 번째 블록에서의 DCT 계수를 의미한다.

## 4. 얼굴 영역의 추출

얼굴 영역의 추출은 앞에서 얻어진 DC 영상에서 이루어진다 본 논문에서는 빠르고 정확한 얼굴 영역의 추출을 위하여 색상 및 형태 정보에 기반한 얼굴 후보 영역 설정 방법과 K-L 변환 및 역변환의 오차를 이용한 얼굴 영역 추출 방법을 제안한다 두 단계의 처리 방법을 사용함으로써 처리 시간을 단축시킬 수 있었으며, 정확도를 향상시킬 수 있었다

### 4.1 얼굴 후보 영역의 추출

얼굴 후보 영역의 추출은 얼굴의 색상 및 형태 정보를 사용한다 색상 정보를 이용한 방법은 MPEG의 색차 신호인  $C_b, C_r$  성분에 기반한 이진화 방법을 사용하였다. 즉 피부 색에 해당하는 색차 신호의 임계값을 적용하여 얼굴 후보 영역을 설정하는 것이다 MPEG의 4.1.1 표준 형식의 경우 4개의 휘도 신호에 대하여 하나의 색차 신호가 형성되므로 설정된 후보 영역은 DC 영상에서 4개의 화소가 기본 단위가 된다 설정된 후보 영역에 대해서 얼굴의 형태 정보에 기반한 얼굴의 크기, 가로세로비의 임계치를 적용하여 이를 크게 넘어서거나 못미치는 경우를 제외하여 최종적인 얼굴 후보 영역을 설정하였다



그림 2. 얼굴 후보 영역 추출의 예

### 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 임의의 영상을 고유 얼굴 영상들로 이루어진 축으로 사영한 후 이를 다시 역변환했을 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 수행하므로서 얼굴의 크기 및 위치를 결정하는 방법을 사용하였다

5. 얼굴 인식

DC 영상으로부터 얼굴 영역이 추출된 프레임에 대하여 GOP 단위의 압축 복원을 수행하여 얻어진 영상에 대하여 얼굴 인식을 수행하였다 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  $I > P > B$ 의 우선 순위를 설정하고 얼굴 영상이 1초내에 다른 인물로 변화하지 않는 가정하에 하나의 GOP 내에서 우선 순위가 높은 하나의 프레임만에 대해 얼굴 인식을 수행한다 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 본 논문에서는 200장의 표본 영상을 선정하고 K-L 변환을 수행하여 상위 40개의 주성분 값을 인식에 사용하였다.

6. 실험 및 성능 분석

6.1 실험 환경

제안된 방법을 검증하기 위한 실험에는 Pentium 166MHz의 PC를 사용하였으며, 352x240의 해상도를 갖는 MPEG-1 형식의 압축 비디오를 사용하였다 실험에 사용된 비디오는 TV 뉴스와 드라마로부터 얻어졌으며 제안된 방법의 성능을 검증할 수 있도록 다양한 크기, 표정, 시점을 갖는 얼굴 영상을 포함하고 있다 그림 3은 실험에 사용된 비디오 데이터의 예이다

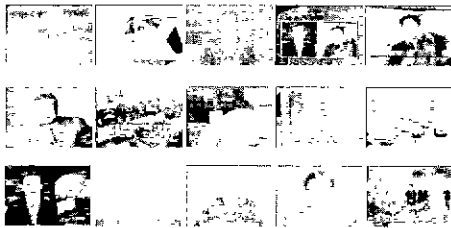


그림 3. 사용된 비디오 데이터의 예

6.2 실험 및 결과 분석

표 1과 표 2는 제안된 방법에 대한 실험 결과이다 얼굴 영역 추출 결과의 경우 단순한 배경과 일정한 크기의 장면 을 바라보는 얼굴 영상에 대해서는 좋은 성능을 보였으나 복잡한 배경하에 얼굴의 시점이 심하게 변화된 영상을 검출해내지 못하였다. 또한 축소된 DC 영상을 사용하여 얼굴 영역을 검정하므로 크기가 작은 얼굴 영상의 경우에는 얼굴 위치에 약간의 오차가 발생하였다

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

제안된 방법은 MPEG 압축 비디오로부터 DC 영상을 구성하여 얼굴 영역을 추출하였으며 얼굴 영역이 추출된 영상에 대한 얼굴 인식을 수행하였다 앞으로 보다 일반화된 얼굴 영역 추출 및 인식을 위해서는 다양한 편진 효과와

여러가지 카메라 동작에 의한 변화를 수용할 수 있도록 비디오 분석 기술의 연구가 함께 진행되어야 할 것이다.

표 1. 얼굴 영역 추출 결과

종류	N	N <sub>C</sub>	N <sub>FP</sub>	N <sub>FP</sub>
뉴스 1	137	102	35	40
뉴스 2	120	91	26	22
드라마	90	71	19	31

표 2. 얼굴 인식 결과

종류	N	N <sub>C</sub>	N <sub>T</sub>
뉴스 1	102	88	14
뉴스 2	91	70	24
드라마	71	65	6

참고 문헌

- [1] J Meng, Y Juan, S F Chang, "Scene Change Detection in a MPEG Compressed Video Sequence" *Proc of SPIE Conf on Digital Video Compression: Algorithms and Technologies*, Vol. 2419, San Jose, pp. 14-25, 1995
- [2] V. Kobola, D S Doermann, 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
- [3] B L Yeo and B Liu "Visual Content Highlighting via Automatic Extraction of Embedded Captions on MPEG Compressed Video," *Proc. of SPIE Conf. on Digital Video Compression Algorithms and Technologies*, Vol. 2668, San Jose, pp. 38-47, 1996.
- [4] H Wang and S F Chang, "A Highly Efficient System for Automatic Face Region Detection in MPEG Video Sequences," *IEEE Trans on Circuits and Systems for Video Technology Special Issue on Multimedia Systems and Technologies*, 1997
- [5] 이 미숙, 이 성환, "얼굴 영상의 이해를 위한 복잡한 배경에서의 정면 얼굴 영역의 실시간 추출," 한국정보과학회 가을 학술발표회 논문집, pp. 609-612, 1996년 10월.
- [6] H A. Rowley, S Baluja, Takeo Kanade, "Nrcal Network-Based Face Detection," *Proc IEEE Computer Society Conf on Computer Vision and Pattern Recognition* pp 203-208 1996
- [7] M. Turk and A Pentland, "Eigenfaces for Recognition" *Journal of Cognitive Neuroscience* Vol 3, No 1, pp 71-86, 1991
- [8] W Konen and S. Eklkehard "ZN-Face: A system for access control using automated face recognition," *Proc. Int Workshop on Automatic Face- and Gesture-Recognition*, Zurich Switzerland, pp 18-23, 1995