

방송영상에서의 등장인물 검색을 위한 고속 얼굴 인식 시스템

정병희⁰ 하명환 김희정 박현선* 이흔진* 김희율*

KBS 기술연구소, 한양대학교 전자통신전과공학과*

{bhjung⁰, mbha, hjkim}@kbs.co.kr, {hspark, hjlee, wykim}@vision.hanyang.ac.kr*

Fast Face Recognition System for Character Retrieval in TV Programs

Byunghee Jung⁰, Myunghwan Ha, Heejung Kim,

Hyunsun Park*, Heunjin Lee*, Whoiyul Kim*

Technical Research Institute KBS, Division of Electrical Engineering Hanyang University*

요약

방송 프로그램이나 영화와 같은 동영상을 인터넷에서 검색하는 서비스가 활성화됨에 따라 특정 인물이 등장하는 부분을 검색하는 기능은 일반 사용자나 프로그램을 만드는 제작자 모두에게 필요한 기능이 되었다. 등장인물 중심의 검색을 위해서는 해당 인물의 얼굴 검출 및 인식 기능이 필수적이며, 특히 방송 영상의 특성에 적합하고 등장인물 검색 서비스에 적용 가능한 얼굴 검출 및 인식 기술이 요구된다. 이를 위해 본 논문에서는 고속 얼굴 인식 시스템을 제안하고, 실시간 수행이 가능한 얼굴 검출 및 인식 알고리즘을 제안한다. 제안한 얼굴 검출 및 인식 알고리즘은 DCT 기법을 전처리 단계로 두어 계산량을 최소화하면서도 특정 값의 정보량은 유지하는 방법을 사용한다. 본 논문에서는 제안하는 알고리즘이 기존 방법에 비해 우수한 성능을 보이며, 실제 방송 영상을 구현된 시스템에 적용하여 시간과 검출률/인식률 측면에서 우수한 결과를 나타낸다.

1. 서론

컴퓨터 소프트웨어, 하드웨어의 발전과 네트워크 성능의 향상으로 인해 방송 프로그램이나 영화와 같은 동영상을 네트워크로 검색해보는 서비스가 활성화되고 있다. 그 중 특정 인물이 등장하는 부분을 검색하는 기능은 일반 사용자나 프로그램을 만드는 제작자 모두에게 유용한 서비스이다. 등장인물 중심의 검색을 위해서는 해당 인물의 얼굴 검출 및 인식 기능이 필수적이며, 특히 방송 영상의 특성에 적합하고 등장인물 검색 서비스에 적용 가능한 얼굴 검출 및 인식 기술이 요구된다.

하루에 방송되는 영상들은 수십 시간 분량이 되므로 고속 얼굴 검출 및 인식 기능이 제공되어야 등장인물 검색 서비스가 가능하다. 방송영상 중 등장인물 중심의 영상들은 대부분 등장인물의 정면 얼굴(두 눈을 포함한 얼굴)을 포함한다. 따라서 요구되는 얼굴 검출 및 인식 기술도 정면 얼굴을 대상으로 한다. 얼굴 검출 측면에서는 정면 얼굴 영역의 실시간 검출과 인식 단계를 위한 눈 영역의 실시간 검출 방안이 요구된다. 얼굴 인식 측면에서는 알려진 얼굴 데이터셋과는 달리 방송에 등장하는 인물들의 얼굴 구성 요소들(예를 들면, 눈의 크기, 위치, 얼굴 형태 등)이 유사하고, 머리 모양 및 기타 배경이 다양하기 때문에 얼굴 내의 특성만으로 변별력을 높이는 빠른 얼굴 인식 방법이 요구된다. 또한 고정된 데이터셋이 아닌 임의의 영상이 입력되는 상황을 고려하여 얼굴 검출/인식 결과를 사전에 확인하고 인물 DB에 등록할 데이터를 필요에 따라 수정할 수 있는 운용시스템도 필수적이다.

본 논문에서는 방송 영상의 등장인물 검색을 위해 각 장면에 있는 인물을 인식하고 각 등장인물의 등장 위치를 분석하여 그 정보를 데이터베이스화할 수 있는 고속 얼굴 인식 시스템을 제안한다. 제안한 시스템은 많은 분량의 영상 데이터도 빠르게 처리할 수 있도록 제안한 고속의 얼굴 검출 알고리즘과 얼굴 인식 알고리즘을 사용한다. 얼굴 검출 알고리즘은 IntegerDCT(Discrete Cosine Transform)를 SVM(Support Vector Machine)의 전처리 단계로 사용하여, 정보를 최소

화하면서도 얼굴 특성 정보는 유지시킨 정보 벡터를 SVM의 입력으로 사용한다. 얼굴 인식 알고리즘은 기존 많이 사용되던 PCA(Principal Component Analysis)/LDA(Linear Discriminant Analysis) 방법에서 PCA에 소용되는 계산량을 최소화한 DCT/LDA 방법을 이용한다.

본 논문은 다음과 같이 구성된다. 2장에서는 제안하는 시스템의 구조와 기능을 설명하며, 3장, 4장에서는 시스템에서 사용하는 얼굴 검출 알고리즘과 얼굴 인식 알고리즘을 제안하고 성능을 분석한다. 5장에서는 구현된 시스템의 인터페이스에 대해 설명하고, 6장에서는 결론과 향후 연구과제에 대해 논한다.

2. 시스템 개요

방송 프로그램은 편성에 의해 동일 요일, 동일 시간에 시리즈로 방송되며, 고정 등장인물이 있는 경우가 일반적이다. 따라서 등장인물 검색을 위해서는 1회 프로그램 등장인물을 먼저 DB에 등록시킨 후 다음 회 프로그램부터는 자동으로 인식하게 하는 것이 효과적인 접근 방법이다.

본 논문에서 제안하는 얼굴 인식 시스템은 그림 1과 같이 초기 얼굴 등록기와 자동 얼굴 인식기 두 부분으로 구성된다. 초기 얼굴 등록기는 얼굴 검출 기능과 검출된 얼굴의 등록 및 수정 기능으로 구성된다. 초기 얼굴 등록기에서는 동영상을 재생하면서 얼굴을 검출한다. 검출된 얼굴이 이미 인물 DB에 등록되어 있으면 등록된 이름을 표시하고, 아니면 unknown으로 표시한다. 검출 결과가 잘못된 경우 사용자는 수동으로 수정하여 정정된 데이터를 인물 DB에 등록할 수 있다. 자동 얼굴 인식기에서는 1회분 혹은 임의영상으로 사전에 등록된 인물 DB를 기반으로 2회분 이후 혹은 임의영상의 얼굴 인식을 수행한다. 인식 결과는 인식 부분을 표시하여 사용자가 인식 결과를 확인할 수 있고, 인물별로 분석하여 등장 구간을 타임라인에 표시해 준다.

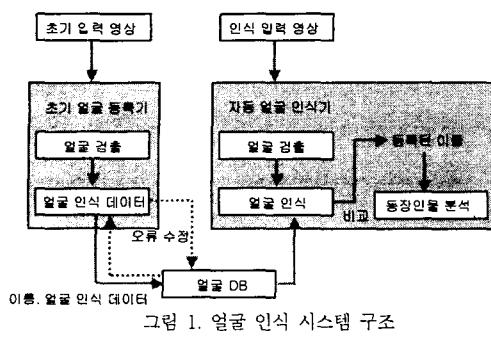


그림 1. 얼굴 인식 시스템 구조

3. 얼굴 검출

얼굴 검출은 특정 기반 검출, 영상 기반 검출로 나눌 수 있다[1]. 특정 기반 검출은 얼굴을 특정 지울 수 있는 눈, 코, 입 등을 먼저 검출하고 그것들의 기하학적 관계를 분석하여 얼굴을 검출하는 방식이다. 영상 기반 검출은 마스크 영상을 신경망이나 SVM과 같은 판별기를 이용하여 얼굴을 판단하는 방법이다. 특정 기반 얼굴 검출은 눈, 코, 입 등을 모두 검출해야 하기 때문에 해상도가 비교적 높아야 하며 알고리즘이 복잡한 단점이 있다. 영상 기반 얼굴 검출은 낮은 해상도에서도 얼굴을 검출할 수 있다는 장점이 있다. 하지만, 영상 기반 얼굴 검출은 마스크 영상의 모든 픽셀을 판별기의 입력 특징 벡터로 사용하고, 다양한 크기의 얼굴을 검출하기 위하여 영상의 크기를 변화시켜가며 전체 영상을 반복적으로 스캔하기 때문에 계산 시간이 오래 걸리는 단점이 있다.

본 시스템에 적용된 제안하는 얼굴 검출 알고리즘은 얼굴 후보 영역을 검출한 후 연산량이 적은 IntegerDCT와 SVM을 이용하여 최종 얼굴 영역을 검출하는 방법이다. 제안하는 얼굴 검출 알고리즘은 24×24 크기의 마스크를 스캔하며 얼굴을 판별한다. 352×240 크기의 영상을 12 스케일로 축소해가고 마스크를 2 픽셀씩 이동하며 스캔하면 총 3340개의 마스크 영상이 나온다. 3340개의 마스크 영상을 SVM으로 검증하는 작업은 시간이 많이 소요되므로 이런 문제점을 극복하기 위하여 IntegerDCT 결과를 SVM의 입력으로 사용한다.

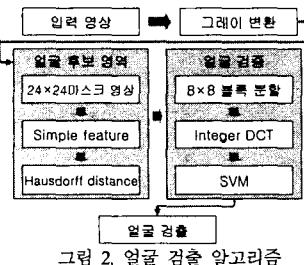


그림 2. 얼굴 검출 알고리즘

3.1 얼굴 후보 영역 검출

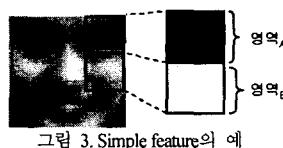


그림 3. Simple feature의 예

Simple feature는 얼굴에 존재하는 어둡고 밝은 영역의 특성을 이용하는 것이다 [2]. 예를 들어 그림 3과 같이 눈 주위는 어둡고 불 주

위는 밝은 특성을 이용하여 얼굴이기 위해서는 (영역B의 픽셀 값의 합)-(영역A의 픽셀 값의 합)이 임계치 이상이 되어야 하는 조건을 만들 수 있다.

본 논문의 실험에서 사용한 Simple feature 들은 그림 4와 같다.

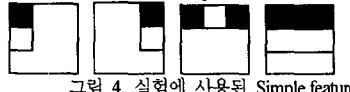


그림 4. 실험에 사용된 Simple feature

다음 단계로 에지 특성으로 에지 템플릿과 마스크 영상의 에지 영상을 구한 후 Hausdorff Distance(HD)를 계산하여 얼굴 후보 영역을 검출한다[3]. 에지 템플릿은 트레이닝 샘플을 소벨 연산을 이용하여 에지 영상으로 만든 후 모든 에지 영상을 중첩한 다음, 중첩된 에지의 개수가 샘플 개수의 2/3 이상이면 1로 그렇지 않으면 0의 값을 갖게 하여 생성한다. 마스크 영상의 에지 또한 소벨 연산을 이용하여 에지의 크기가 임계치 이상이면 1로 그렇지 않으면 0으로 한다.

3.2 최종 얼굴 영역 검출

제안하는 얼굴 검출 알고리즘에서는 얼굴 영역을 최종검증하기 위해 SVM을 사용한다. 먼저 24×24 크기의 영상은 얼굴 검증을 위하여 9개의 8×8 크기의 블록으로 분할된다. 각각의 블록에 대하여 IntegerDCT를 수행하고 SVM의 입력 특징 벡터를 만들기 위하여 DCT 계수의 DC 성분을 제외한 12개의 저주파 DCT 계수를 지그재그 스캔을 이용하여 추출한다. 이때 DC성분을 제외한 이유는 조명 성분을 없애기 위해서다. 최종 추출된 특징 벡터는 9×12=108 차원이 된다.

$$F(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos \left[\frac{(2x+1)u\pi}{2N} \right]$$

$$\text{이 때, } \alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & u = 0 \\ \sqrt{\frac{2}{N}} & u = 1, 2, \dots, N-1 \end{cases} \quad (1)$$

DCT는 압축 성능이 뛰어난 장점이 있지만 식 (1)과 같이 많은 실수 연산이 필요하다. 이런 이유로 본 논문에서 제안한 고속 얼굴 검출 알고리즘에는 부적합하다. Integer DCT는 DCT의 실수 연산을 근사화 하여 덧셈과 쉬프트 연산만을 이용하여 DCT를 계산하는 매우 빠른 알고리즘이다[4].

3.3 얼굴 검출 알고리즘 성능 평가

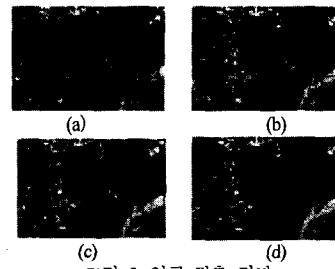


그림 5. 얼굴 검출 결과

실험에 쓰인 데이터는 352×240 크기의 방송 파일이며, PC는 Pentium-IV 1.7 GHz를 이용하였다. SVM 학습을 위한 샘플은 Bootstrapping 알고리즘을 이용하여 구하였으며, 구해진 얼굴 샘플의 개수는 295개이고 비 얼굴 샘플의 개수는 3736개이다. 그림 5(a)는 얼굴 후보 영역이며, 그림 5(b)는 HD를 이용하여 검출된 얼굴 후보 영역이다. 그림 5(c)는 Simple feature와 HD를 순차 적용하여 검출된 얼굴 후보 영역이다. Simple feature와 HD를 이용한 얼굴 후보 영역

검출의 성능이 우수함을 알 수 있다. 그림 5(d)는 최종적인 얼굴 검출 결과이다. 계산 시간은 평균 62 ms/f으로 실시간으로 검출됨을 알 수 있었다.

4. 얼굴 인식

얼굴 인식의 성능을 높이기 위한 기존 연구들은 기하학적인 특징 값 기반 방법과 템플릿 기반 방법, 모델 기반의 방법 등의 연구가 있다. 대표적인 얼굴 인식 기법으로는 PCA, LDA, NN(Neural Network), SVM, HMM (Hidden Markov Model) 등이 있다[5]. 제안하는 얼굴 인식 알고리즘은 LDA에 기반 한다.

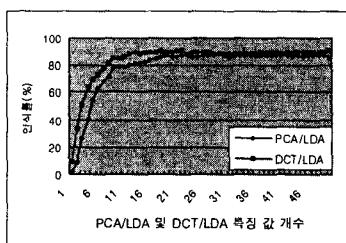
4.1 DCT/LDA 얼굴 인식

LDA는 얼굴 영상들을 가장 잘 분리 할 수 있는 평면으로 투영시켜서 얼굴을 인식하는 방법으로써 보통 PCA를 병합한 PCA/LDA 방법을 이용한다[6]. PCA/LDA 방법은 높은 인식률을 보이지만 PCA에서 기저 벡터를 계산하는데 많은 계산량을 필요로 하고 PCA의 기저 벡터는 트레이닝 셋의 변화에 따라 변화하는 단점이 있다. 본 논문에서는 이런 문제점을 개선하기 위하여 DCT/LDA를 제안한다. PCA 대신 DCT를 이용하게 되면 기저 벡터를 계산하는 트레이닝 과정 없이 DCT 계수를 바로 추출할 수 있다. 또한, DCT는 트레이닝 셋에 관계없이 일정한 기저 벡터를 가지는 장점이 있다[7].

DCT를 수행하기 전에 조명의 변화에 대처하기 위해서 얼굴 영상에 ZMST (Zero Mean Unit Variance)를 수행한다[8]. 이 ZMST는 영상의 밝기 값에 영상의 평균을 빼고 표준편차로 나눠서 밝기 성분을 정규화한다. 전 처리가 끝나면 얼굴 영상 전체에 대해서 DCT를 수행하여 DCT 계수를 추출한다. 조명의 변화를 감소시키기 위해서 DC 값은 사용하지 않으며 저주파 영역을 래스터 스캔하여 계수 값들을 추출한다. 이렇게 추출된 DCT 계수를 가지고 LDA를 트레이닝하여 LDA의 기저 벡터를 추출한다. 추출된 기저 벡터를 통하여 얼굴 영상들의 DCT/LDA 특징 값을 추출한다. 얼굴 인식은 추출된 DCT/LDA 계수들 간에 가중치가 있는 유클리디언 거리를 사용하며 거리가 가장 적은 얼굴을 그 사람의 얼굴로 판단한다[6].

4.2 성능 평가

제안한 얼굴 인식 알고리즘의 성능 검증을 위해 PCA/LDA와 비교하였다. 실험 영상은 MPEG-7 얼굴 인식 기술자의 실험에서 사용되었던 데이터 셋 중 900장의 Alticom 데이터 셋과 3270 장의 MPEG-7 데이터 셋을 사용하였다. 실험 영상에 사용된 영상은 46×56 크기의 그레이 영상으로써 Alticom 데이터 셋은 사람 당 9장, MPEG-7 데이터 셋은 사람 당 5장으로 구성하였다. 각 영상들은 조명과 포즈의 변화가 있고 눈의 위치에 대해서 정규화되어 있다.



벡터를 생성함을 알 수 있다.

5. 시스템 구현

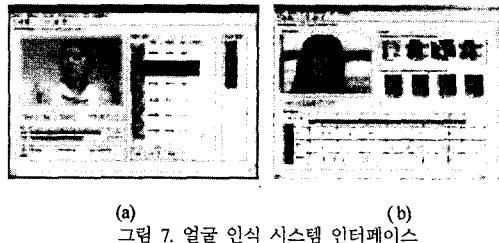


그림 7. 얼굴 인식 시스템 인터페이스

시스템은 VisualC++6.0을 사용하여 구현하였다. 그림 7(a)와 그림 7(b)는 드라마 “가을동화” 영상을 얼굴 검출/인식한 결과를 보여준다. 그림 7(a)에서는 영상 재생과 동시에 검출되는 얼굴을 리스트박스로 보여주고 해당하는 인물의 DB에 등록된 얼굴 이미지도 보여준다. 그림 7(b)에서는 등록된 인물을 기반으로 새로운 입력 영상에 대해 재생과 동시에 인식된 얼굴을 보여주며, 전체 영상에서의 각 인물별 등장 구간을 타임라인에 표시하고 있다.

6. 결론 및 향후 연구과제

본 논문에서는 대용량 방송 영상에서 등장인물 검색 기능에 적용 가능한 고속 얼굴 인식 시스템을 제안하고 구현하였다. 고속 얼굴 인식을 위하여 본 논문에서는 개선된 얼굴 검출 및 인식 알고리즘도 제안하였다. 얼굴 검출 알고리즘은 검출된 얼굴 후보 영역에 계산량이 축소된 IntegerDCT 결과를 SVM 입력으로 적용하여 얼굴 검증 과정을 수행한다. 제안한 얼굴 검출 알고리즘은 검출률은 높이고 실시간 수행시간을 보장한다. 얼굴 인식 알고리즘은 기존의 PCA/LDA 방법에 비해 계산량은 줄이고 트레이닝 셋에 관계없이 일정한 기저 벡터를 가지는 DCT/LDA 방법을 사용한다. 제안한 얼굴 인식 알고리즘 역시 실시간성을 보장하고 적은 트레이닝 셋에도 성능이 저하되지 않는다.

향후 연구과제로는 인식률을 더욱 향상하기 위한 얼굴 내의 구성 요소별 인식 과정을 추가하는 것과 등장인물 검색 기능을 실 인터넷 검색 환경에서 제공하기 위해 적절한 절의 시스템 모델을 설정하는 것 등이 있다.

7. 참고 문헌

- [1] Min-Hsuan Yang, David Kriegman, and Narendra Ahuja, Detecting Faces in Images: A Survey, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, 2001.
- [2] P. Viola and M. Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, *Proceedings IEEE Conf. on CVPR* 2001.
- [3] O. Jezorsky, K. Kirchberg and R. Frischholz, “Robust Face Detection Using the Hausdorff Distance”, *Int. Conf. on AVBPA*, Halmstad, Sweden, Springer, pp. 90-95, 2001.
- [4] Ying-Jui Chen, Soontorn Oraintara, and Truong Nguyen, Integer Discrete Cosine Transform (IntDCT), *Int. Conf. on Information, Communications and Signal Processing*, Singapore, Dec. 1999.
- [5] R. Chellappa, C. L. Wilson, S. Sirohey, *Human and Machine Recognition of Faces: A Survey*, *Proceedings of the IEEE*, Vol. 83, No. 5, pp. 705-741, 1995.
- [6] W. Zhao, R. Chellappa, and A. Krishnaswamy, “Discriminant analysis of principal components for face recognition”, *In Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition*, pp. 336-341, 1998.
- [7] Ziad M.Hafed and Martin D. Levine, “Face Recognition using the discrete cosine transform”, *International Journal of Computer Vision*, Vol. 43, No.3, pp.167-188, July 2001.
- [8] Yongping Li, “Linear Discriminant Analysis and its application to Face Identification”, *Ph. D. Thesis Paper*, University of Surrey, September 2000.