

‘얼굴에 의한 질의’ 환경을 위한 검색 시스템

고병철^{*}, 안준한, 이해성, 변혜란
연세대학교 컴퓨터 과학과

Retrieval System for Query-by-face environment

ByongChul Ko^{*}, Jun-Han Ahn, Hae-Sung, Lee, Hyeran Byun
Dept. of Computer Science, Yonsei University

요 약

최근 몇 년 동안 멀티미디어 정보의 급격한 증가와 더불어, 사용자에게 다양한 형태의 검색 환경 제공하기 위한 연구들이 꾸준히 이루어지고 있지만, ‘얼굴에 의한 질의(Query-by-Face)’에 대한 연구는 다른 검색 방법에 비해 상대적으로 미약한 편이다. 얼굴 검색은 일반적인 내용기반 검색 분야에 비해 어려운 분야로 현재 대부분의 얼굴 인식 및 검색 시스템들은 일정한 배경 및 조명, 동일한 얼굴 크기를 갖는 한정된 형태의 데이터만을 사용한다. 본 논문에서는 영화 비디오로부터 추출된 대표 프레임 중 사용자가 원하는 등장 인물을 찾기 위한 기술 개발을 위해 비디오 영상으로부터 얼굴을 검출하고 인식하는 방법을 제안한다. 기존의 방법들은 실험 영상이 제한되어 있거나 인식의 정확성을 위해 몇 개의 부수적인 얼굴 데이터를 별도로 보관해야만 했지만 본 논문에서는 배경과 두드러진 특징을 갖는 얼굴 색을 이용하여 얼굴 영역 검출 속도를 향상시키고, 웨이블릿 변환과 하우스도르프 거리(Hausdorff distance)를 이용하여 별도의 데이터가 필요 없이 얼굴을 인식이 가능한 시스템을 설계하였다. 또한, 영화 비디오 및 뉴스, 인터뷰 비디오 등 다양한 형태의 배경 및 조명, 크기 변화를 갖는 데이터에 대한 실험결과를 통해 본 논문에서 제안하는 방법에 대한 성능 평가 실시하였다.

1. 서 론

최근 몇 년 동안 멀티미디어 정보의 급격한 증가와 더불어, 사용자에게 다양한 형태의 검색 환경 제공하기 위한 연구들이 꾸준히 이루어지고 있다. 대부분의 내용기반 검색(Content-Based Image Retrieval) 환경들은 영상에 의한 질의(Query-by-image)[1][2], 스케치에 의한 질의(Query-by-sketch)[2], 영역에 의한 질의(Query-by-region)[3]들과 같은 검색 환경을 제공하고 있지만, 얼굴에 의한 질의에 대한 연구는 상대적으로 미약하다. 얼굴 검색은 일반적인 내용기반 검색 분야에 비해 어려운 분야로 현재 대부분의 얼굴 인식 및 검색 시스템들은 일정한 배경 및 조명, 동일한 얼굴 크기를 갖는 한정된 형태의 데이터만을 사용한다. 얼굴 검출 및 인식은 주로 범인 인식 시스템으로부터 각종 보안 시스템, 신용카드, 그리고 디지털 영화 편집 시스템과 같은 다양한 분야에 적용되어 오고 있다. 본 논문에서는 디지털 비디오로부터 보다 효율적이고 의미 있는 내용기반 검색(CBIR)을 위해 비디오를 파싱(Parsing)하고 추출된 대표 프레임(Key frame)으로부터 특정 인물을 자동적으로 인식 및 검색할 수 있는 시스템을 설계하였다. 본 논문에서는 비디오 샷 경계면 분할 및 대표 프레임 추출을 위하여 [4]에서 제안한 하이브리드 비교법과 대표 프레임 추출 알고리즘을 사용하였다. 우선 비디오를 각 샷으로 분할 한 뒤, 각 샷으로부터 대표 프레임을 선출하게 된다. 이렇게 선

출된 대표 프레임을 이용하여 시스템은 질의하는 인물과 가장 유사성을 지닌 대표 프레임을 검색 결과로 보여 줌으로서 사용자는 영화 전체를 일일이 검색하지 않더라도 원하는 인물이 출현하는 장면을 검색 할 수 있게 된다. 본 논문의 구성은 다음과 같다. 2장에서는 얼굴 검출 및 인식에 대한 기존 연구를 소개하고, 3장에서는 본 논문의 알고리즘을 소개한다. 마지막으로 4장에서 일반 영화와 뉴스 비디오에 대하여 실험한 결과를 보여 주고, 5장에서 결론 및 향후 개선방향을 제시한다.

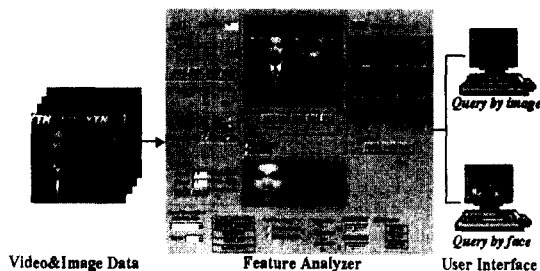


그림 1. 검색 시스템 개략도

2. 얼굴 검출 및 인식에 대한 기존 연구 방법

얼굴 검출 및 인식에 대한 기존의 연구 방향은 크게 영상으로부터 얼굴 영역만을 검출해 내는 얼굴 영역 검출 기법과 검출된 얼굴 영상으로부터 데이터베이스에 있는 얼굴 데이터와의 유사성 측정을 통해 동일한 얼굴을 찾아내는 얼굴 인식 분야로 나뉘어 진다. 우선 얼굴 영역 검출에 대한 연구를 살펴보면 얼굴 형판에 기반한 방법[5]과 얼굴 요소에 의한 방법[6], 최근에 가장 많이 사용되는 얼굴 색에 의한 검출 방법등이 있다. 이 방법[7]은 일반적으로 얼굴 색이 다른 배경 색에 비해 두드러진 특징을 갖는다는 사실을 이용하여 얼굴 영역을 검출해 낸다. 다음으로 얼굴 인식에 대한 연구는 아래와 같이 대표적으로 몇 가지로 나눌 수 있다.

① 주성분(PCA) 분석 방법

Turk 와 Pentland[7]가 제안한 주성분 방법은 얼굴 인식 분야에서 가장 일반적으로 사용되는 방법이다. 이 방법은 얼굴 표정, 각도, 조명등에 둔감하지만, 얼굴 크기의 변화에 매우 민감하고, 계산량과 추가적인 저장공간을 필요로 한다.

② Gabor wavelet 방법

Gabor wavelet[8]을 이용한 방법의 경우 얼굴의 크기, 각도, 표정 등에 무관한 인식이 가능하지만, 한 영상을 인식하기 위한 처리 시간이 많이 소모되어 대용량의 데이터베이스를 대상으로 하는 얼굴 영상 검색에는 부적합하다.

3. 색상 기반 얼굴 영역 검출 및 인식방법

3.1 색상에 의한 후보 얼굴 영역 검출

영상에서 얼굴 영역은 다른 배경 색과 확연히 구분되는 특징을 가지고 있을 뿐만 아니라, 색상에 의한 얼굴 검출 방법은 계산 시간이 빠른 장점을 가지고 있다[10]. 따라서 이러한 특징을 이용하여 우선적으로 영상으로부터 얼굴 후보 영역을 찾아낸다. 이때, 명암 및 피부색에 의한 오류를 막기 위해 RGB 색상을 HSV로 변환시키고 H와 S값에 대해 식(1)을 적용하여 얼굴 후보 영역을 선출한다. 이것은 HSV 색상 영역이 인간의 시각과 유사한 특징을 가지고 있으므로 RGB영역에 비해 좀더 정확한 얼굴 영역을 검출해 낼 수 있기 때문이다[7].

$$FR(x, y) = 1 \text{ if } H \leq T_1 \cap S \leq T_2 \quad (1)$$

0 otherwise

FR: Face Region, H: Hue, S: Saturation

다음으로 식(1)에 의해 추출된 1차 얼굴 후보 영역에 대하여 이진 모폴로지(Binary morphology)를 적용한다. 우선 후보 얼굴영역 안쪽에 눈이나 입 색깔로 인해 생긴 구멍을 채워 주기 위하여 팽창(Dilation)을 적용하고, 얼굴과 유사한 색을 가지는 조그만 물체나 노이즈를 제거하기 위하여 녹임(Erosion)을 적용한다. 이렇게 해서 남은 후보 얼굴 영역에 대하여 가로, 세로 스캐닝(Horizontal & Vertical Scanning)을 실시한다. 이것은 얼굴 영역이 다른 배경이나 잡영(noise)에 비해 크다는 가정에 기초한다. 스캐닝 작업은 우선 가로축으로부터 시작되어, 같은 값을 갖는 화소의 수가 연속적으로 일정 수 이상 될 경우, 해당 가로축의 한 영역에 포함되는 화소일 가능성이 크므로 여기에 해당하는 화소들을 모두 1로 만

들어준다. 이 작업을 가로축 각 화소마다 적용한다. 가로축 작업이 끝나면 다시 세로축에 이 방법을 적용한다. 1차적으로 추출된 후보 얼굴 영역의 경우 인식을 위한 영상으로는 부적합하므로 보다 정확한 얼굴 영역 추출을 위해서 1차 후보 얼굴 영역에 대해 해당 영역의 RGB 컬러 성분을 YCbCr로 변환하여 위의 과정을 반복한다. 마지막으로 추출된 얼굴 영역에 대하여 Y축 투영을 실시하여 정해진 임계값보다 작은 영역에 대해서는 후보 영역에서 제거시킨다. 이것은 보다 정확한 얼굴 영역을 추출하기 위해 목 부분을 제거시키기 위한 작업이다. 그런 뒤에 최종적인 얼굴영역의 가로축과 세로축의 길이를 측정하여 세로축의 길이가 가로축보다 2배 이상 클 경우 세로축의 길이를 가로축과 1:1.5 비율로 잘라준다. 지금까지의 얼굴 영역 추출 알고리즘을 요약하면 그림3과 같다.

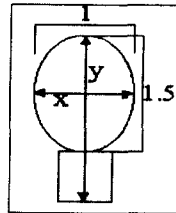


그림 2.
얼굴 영역
중형비

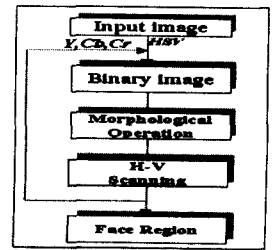


그림 3. 얼굴 영역 추출
개념도

3.2 얼굴 인식 방법

추출된 얼굴 영역은 촬영된 카메라의 거리에 따라 서로 다른 크기를 가지므로 보다 정확한 인식을 위해서는 얼굴 영역을 일정 크기로 정규화 시켜주는 작업이 필요하다. 즉, 얼굴영역이 작을 경우 그 크기에 따라 2배 혹은 3배 확대(Interpolation) 시켜 줌으로써 얼굴 영역을 일정하게(160x160) 유지하도록 하였다. 그런 뒤에 얼굴 영역에 포함되어 있는 노이즈를 제거하기 위해 가우시안 필터를 적용하였다.

다음으로는 가우시안 필터를 통과시킨 영역에 대해 웨이블릿 변환을 가한다. 웨이블릿 변환은 주성분 분석(PCA)에 비해 처리 속도가 빠르고, 추가적인 저장 공간을 필요로 하지 않는다. 본 논문에서는 Biorthogonal wavelet frame구조의 3단계 웨이블릿 변환에 의한 저주파 성분만을 이용하여 1차 특징값으로 사용하였다. 이때 질의 영상과 데이터베이스 영상간의 유사성 측정은 식(2)를 이용하였다.

$$C_{L1}(x, y) = 1 - \frac{\sum_i |x_i - y_i|}{\sum_i (|x_i| + |y_i|)} \quad (2)$$

C_{L1} : 유사성 측정치(L_1 norm)

x : 입력 영상, y : 데이터베이스 영상

웨이블릿 변환을 이용한 특징값으로부터 대응되는 화소간의 유사성을 측정된 후에 다시 얼굴 영역으로부터 Hsieh[9]등이 제안한 경계선 검출 알고리즘을 사용하여 경계선(edge)을 검출한다. 이렇게 검출된 경계선에 대해서 하우스도르프 거리(Hausdorff Distance)를 측정한다. 하우스도르프 거리는 물체간에 정확한 겹침보다는 근사값을

측정하므로 모양이 일정하지 않은 물체의 비교에 있어 보다 유연한 측정을 할 수 있다[10].

하지만 일반적인 하우스도르프 거리는 잡영에 민감하므로 본 논문에서는 MHD(Modified Hausdorff Distance)를 사용하였다. 이 경우 각 집합들에서 점들간의 거리는 정규화 되어 나타나게 되므로 잡영에 덜 민감한 결과를 나타낼 수 있다.

$$\text{집합 } A = \{a_1, \dots, a_p\}$$

$$\text{집합 } B = \{b_1, \dots, b_p\}$$

$$h(A, B) = \frac{1}{N_a} \sum_{a \in A} \min_{b \in B} \|a - b\|$$

where ,

$N_a = p$: 집합 A에 포함되는 화소수

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (3)$$

이때, 한 점에 대해서 전체 얼굴 영역을 탐색 할 경우 상당한 시간이 필요하므로 본 논문에서는 질의 영상과 데이터베이스 영상 사이에 일정한 탐색 블록(16x16)을 정하여 약 262.5배의 비교 횟수를 줄일 수 있었다.



그림 4. 검출된 얼굴 영역에 대한 3단계 lowpass영역(좌)과 경계선 영상에 대한 탐색 영역(우)

3.3 결정 트리틀 이용한 유사 얼굴 영상 검색

식(2)과 식(3)에서 얻어진 2개의 거리값을 이용하여 최종적으로 질의 영상과 데이터베이스 영상간의 유사성을 측정한다. 유사성 측정을 위해서 본 논문에서는 결정 트리(Decision Tree) 방법을 사용하였다. 이것은 근 노드로부터 시작하여, 여러 개의 가지 노드(Branch node)를 생성하고 조건을 만족하는 최소 노드를 결과로 선택하게 된다.

$$\text{Score} = E + (1 - H) \cdot w; \quad (4)$$

E : Euclidean Distance, H : Hausdorff Distance

w : Penalty weight

4. 실험 및 성능 분석 및 인식방법

본 논문에서는 실험 데이터로 뉴스 비디오에서 추출된 앵커 샷의 대표 프레임으로부터 총 51명의 360개 영상 데이터와 일반 영화에서 추출된 100개의 프레임, 인터뷰 샷에서 추출된 55개의 다양한 영상 데이터를 사용하였고 황인, 백인이 포함된 영상에 대해서만 실험하였다. 우선 총 20명에 대해 무작위 질의를 하고, 검색 결과중에서 상위 3.5개의 결과에 대해서 precision을 이용하여 정확도를 측정하였다 [표1].

5. 결론 및 향후 개선 방향

본 논문에서는 영화 비디오로부터 추출된 대표 프레임 중 사용자가 원하는 등장 인물을 찾기 위한 기술 개발을 위해 비디오 영상으로부터 얼굴을 검출하고 인식하는 방법을 제안하였다. 기존의

	Precision	
검색 프레임수	k=3	k=5
검색 정확율	94.5	82

표 1. 검색 정확율

방법들은 실험 영상이 제한되어 있거나 데이터베이스 영상이 정확성을 위해 몇 개의 부수적인 데이터를 보관해야 하므로 저장 공간의 비효율성 지적되어 왔다. 또한 얼굴 검출 및 인식을 위한 시간도 문제점이었다. 하지만 본 논문에서 제안하는 방법은 얼굴 검출 및 인식 시간을 크게 단축시킬 수 있었고, 또한 별도의 부수적인 데이터베이스를 필요로 하지 않는다. 하지만, 보다 정확한 실험을 위한 보다 많은 수의 데이터베이스가 필요하고, 다양한 크기와 기울어짐을 갖는 얼굴들에 대해서도 정확한 결과를 보여줄 수 있는 알고리즘이 필요하다.



그림 5. 검색 결과 (상위 5개 영상)

6. 참고문헌

- [1] C. Faloutsos, M. Flickner, W. Niblack, D. Petkovic, W. Equitz, R. Barber, *Efficient and Effective Querying by Image Content*, Research Report #RJ 9203 (81511), IBM Almaden Research Center, San Jose, Aug. 1993.
- [2] M.J. Swain, *Interactive indexing into image database*, Proc. SPIE: Storage and Retrieval Image and Video Database 1908, February 1993, 95-103.
- [3] J.R. Smith and S.F. Chang, *VisualSEEK: A Fully Automated Content-Based Image Query System*, ACM Multimedia 1996, Boston MA, Nov, 1996.
- [4] 고병철, 변해란, "비디오 샷 경계면 분할기법 비교를 통한 대표 프레임 추출" '98 정보과학회 가을 학술발표 논문집(II) ,pp 512 - 514
- [5] 고재필, 변해란, "고유얼굴 기반의 얼굴 형판을 이용한 얼굴 영역 추출" 1998. 6. 연세대학교 석사 졸업논문
- [6] Surendra R, Krishnamurthy A, "Face Recognition using Transform features and Neural Networks", *Pattern Recognition*, Vol. 30, No. 10, pp. 1615-1622, 1997
- [7] N. Herodotou, K.N. Plataniotis, and A.N. Venetsanopoulos, "A Content-Based Storage and Retrieval Scheme for Image and Video Databases", SPIE Visual communication and Image Processing, pp 697-708, 1998.
- [8] M. Turk and A. Pentland, "Eigenfaces for recognition, *J.Cognitive Neuroscience*. 3(1), 71 - 86(1991)
- [9] 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*, pp.18-23, 1995.
- [10] Douglas Chai and King N. Ngan, "Locating Facial Region of a Head-and-Shoulders Color Image" *International conference on Automatic Face and Gesture Recognition*, pp. 124 - 129. 1998. 4
- [11] Jun-Wei Hsieh, Hong-Yuan Mark Liao, Ming-Tat Ko, and Kuo-Chin Fan, "A New Wavelet-Based Edge Detector via Constrained Optimization", *Image and Vision Computing*, Vol. 15, ISS.7, pp. 511-527, 1997.