

제스처를 이용한 새로운 내용기반 영상 검색 기법

고병철⁰ 변혜란
연세대학교 컴퓨터과학과
(soccer1, hrbyun}@aipiri.yonsei.ac.kr

A New Content-Based Image Retrieval Scheme: 'Query-by-Gesture'

ByoungChul Ko⁰ Hyeran Byun
Dept. of Computer Science, Yonsei University

요 약

본 논문에서는 새로운 내용기반 영상 검색 방법인 'query-by-gesture'를 제안하고 이를 본 논문의 영역기반 영상 검색 도구인 FRIP시스템에 적용하였다. 'query-by-gesture' 검색 방법을 이용하여, 사용자는 마우스나 다른 스케치 도구를 사용하지 않더라도 컴퓨터에 부착된 카메라를 이용하여 쉽고 편리한 방법으로 찾고자 하는 객체를 검색할 수 있다. 또한 본 논문에서 제안하는 'query-by-gesture' 방법은 다른 동작 인식 방법에서 문제점으로 제기되는 속도 문제를 해결하기 위해 색상을 이용하여 손 영역을 찾아내고 찾아진 손가락 끝점에 local 윈도우를 적용시켜 빠르고 효율적인 검색 환경을 제공하도록 설계 되었다.

1. 서론

과거 수년 동안 내용기반 영상검색(CBIR)에 대한 연구는 상당한 발전을 가져왔다. 더불어, 최근 몇 년 동안에는 텍스트 기반의 초기 검색 방법의 한계를 극복하기 위해, 'query-by-image[1][5]', 'query-by-sketch[1][3][4][5]', 'query-by-object[1][2][3]'와 같은 다양한 검색 환경을 제공하는 내용기반 영상검색에 대한 연구가 진행되어 왔다. 하지만 현재 대부분의 영상검색 방법은 'query-by-image'가 가지는 단점을 보완하기 위하여 스케치 툴을 이용하는 'query-by-sketch'와 객체단위 또는 영역단위 검색 방법인 'query-by-object'에 초점이 맞춰지고 있다. 'query-by-sketch' 기법은 사용자가 검색하고자 하는 영상을 직접 스케치 함으로써 보다 의미 있고 정확한 검색환경을 제공하지만 다음과 같은 단점을 가지고 있다.

- 대부분의 스케치 도구는 많은 기능을 가지고 있음으로 영상처리에 대한 기본적 지식이 없는 초보자가 사용하기에는 부적합하다.
 - 마우스만을 이용하여 정확한 객체를 그리는 것은 쉽지 않다.
- 따라서, 이러한 단점들을 극복하고, 사용자에게

보다 편리한 검색 환경을 제공하기 위하여 본 논문에서는 본 논문의 영역기반 영상검색 도구인 FRIP 시스템을 이용하여 인간의 손 동작(gesture)을 인식하고 이를 통해 유사한 영상 또는 객체를 검색하는 새로운 개념의 영상 검색 방법을 제안한다.

2. 관련연구

2.1 손 동작 인식

인간의 손 동작은 HCI(Human-Computer Interface)분야에서 키보드, 마우스, 터치 패널(panel)을 대신하는 중요한 역할을 한다. 초기 연구에서는 Glove-based 시스템을 이용한 연구가 대부분이었지만, 이것은 장갑(glove)을 끼어야 하는 단점 때문에 최근의 연구에서는 스테레오 혹은 여러대의 카메라를 이용하여 손동작을 인식하도록 하고 있다. 또한, 이러한 손동작 인식은 VR환경이나, TV 리모콘, CAD와 같은 컴퓨터 디자인, 컴퓨터 게임등에 응용 될 수 있다. 하지만 컴퓨터 비전에 기반한 효율적인 HCI환경을 제공하기 위해서는 다음과 같은 조건이 만족 되어야 한다.

- 반응시간이 빨라야 한다.
- 알고리즘은 신뢰성이 있어야 한다. 사람이 달라지거나 배경이 달라지더라도 결과는 같아야 한다.
- 장비의 가격이 비싸지 않아야 한다.

본 논문에서는 위 3가지 조건을 고려하여 본 논문의 영역기반 영상검색 도구인 FRIP(Finding Region In the Pictures)시스템에 손가락 동작 인식을 영상검색의 질의 환경으로 적용시켰다.

FRIP시스템은 영역 기반 영상 검색 도구로서, 데이터베이스내의 영상을 분할하고 각 영상내의 객체 또는 영역에 대해 색인을 구축하여 저장한다. 검색시에는 질의 영상을 역시 분할하고 사용자가 검색하고자 하는 영상을 선택하면 데이터베이스내의 색인값과 비교를 통해 결과 영상을 보여준다. 색인을 위해 FRIP시스템에서는 색상(Color), 질감(Texture), 크기(Scale), 모양(Shape), 위치(Location)와 같은 5개의 특징값을 추출한 후 가우시안 정규화(Gaussian normalization)에 의해 데이터베이스에 색인으로 저장한다. FRIP 시스템에 대한 자세한 내용은 [6]에서술되어 있다.

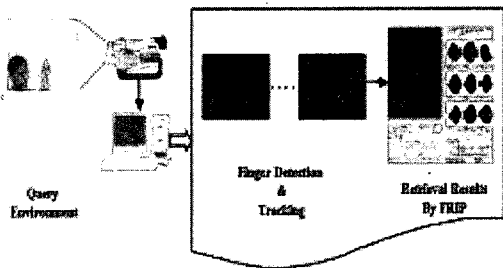


그림 1. 'Query-by-Gesture'를 위한 FRIP 시스템 구성

2. 손가락 끝점(fingertip) 탐색과 추적(tracking)

2.1 손 영역 탐색

본 시스템에서는 손동작 중 손가락 끝점을 탐지하고 추적하여 사용자가 찾고자 하는 객체를 인식하도록 하였다. 우선, 카메라로부터 사용자의 손동작을 인식 받고 색상을 이용하여 손 영역을 배경과 분리해낸다. 이때, 손 영역을 분리해 내기 위해 본 시스템에서는 HSV 색상 중 V(value)를 제외한 H(Hue)와 S(Saturation)만을 이용하였다(식 1).

$$HR(x, y) = \begin{cases} 1 & 340^\circ \leq H \leq 360^\circ \text{ and } S \geq 20\% \\ 1 & 0^\circ \leq H \leq 60^\circ \text{ and } S \geq 20\% \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

HR: 후보 손 영역

식(1)로 찾아진 손 후보 영역에서 유사한 색을 가지는 다른 객체 또는 잡음을 제거하고, 영역 내에 포함된 눈, 입과 같이 다른 색을 가짐으로써 나타나는 구멍(hole)으로 등을 제거해 주기 위해, 메디안 필터(median filter)와 모폴로지 닫힘(morphological closing)을 적용시킨다.

2.2 손가락 끝점 탐색 및 추적

후보 손 영역을 찾아낸 후에 실제 손 영역을 구분해 내어야 한다. 본 시스템에서는 실제 손 영역은 후보

손 영역 중에서 가장 움직임이 많은 영역으로 가정하고 첫 번째 프레임에서 영역들의 중심점과 두 번째 프레임에서 영역들의 중심점 간에 거리를 계산하여 거리가 가장 먼 영역을 손 영역으로 간주 하였다.

$$H_C = \max \left(\sum_{n=1}^R |C_{i-1}^i - C_i^i| \right) \quad (2)$$

H_C : 실제 손 영역의 중심점

C_{i-1}^i : 첫 번째 프레임의 i 번째 영역의 중심점

C_i^i : 두 번째 프레임의 i 번째 영역의 중심점

R: 영역의 총수

식(2)를 통해 실제 손 영역을 찾은 뒤에, 손가락 끝점은 손 영역의 중심점으로부터 가장 먼 거리의 점으로 가정하였다. 다음으로는 연속되는 다음 프레임에서 손가락 끝점 탐지 속도를 높이기 위해, 지역 탐색 창(local search window)를 손가락 끝점을 중심으로 적용시킨다. 그리고 이전 프레임의 손끝 좌표와 현재 프레임의 손끝 좌표가 원도우 크기의 반을 초과할 경우 새로운 손끝 좌표는 이전 프레임의 손끝 좌표를 유지한다. 이것은 사용자의 손가락이 급격히 이동하였을 경우 카메라가 캡처하지 못하는 오류를 방지하고 다른 영역들의 움직임과 구분하기 위해서이다. 손가락 끝점을 이용한 질의 영상구성과정은 그림 2와 같다.

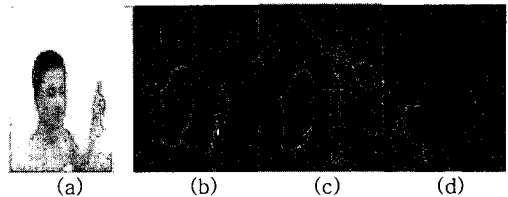


그림 2. 손가락 이용한 질의 객체 생성 (a) 입력 영상 (b) 손가락 끝점 검출 (c) 지역 탐색 윈도우를 이용한 객체 생성 (d) 질의 객체

3. 질의 객체 특징 추출

질의 객체가 완성된 후에, 객체로부터 유사성 측정을 위한 특징 값들을 추출한다. 객체로부터 추출되는 특징 값은 아래와 같다.

(1)색상: 사용자 인터페이스로부터 RGB 3개의 색상을 선택 한다.

(2)질감: 'query-by-gesture' 환경에서는 질감 성분은 제외시킨다.

(3)크기: 객체를 화소 값으로 채운 뒤에 총 화소의 수를 크기 정보로 사용한다.

(4)위치: 영역의 중심좌표 값을 위치 정보로 사용한다.

(5)모양: 본 논문에서 제안하는 변형 시그네춰(Modified Radius-based signature)를 이용하여 12개의 반지름 값을 모양 정보로 사용한다.

최종적으로 질의 객체로부터 추출된 9개의 특징 값은 다음과 같다.

Object

keys: ObjectNo

{

```

attribute int AverRed(AR);
attribute int AverGreen(AG);
attribute int AverageBlue(AB);
attribute int NumberOfPixels(NP);
attribute int CenterOfX(Cx);
attribute int CenterOfY(Cy);
attribute int MajorLength(Rmax);
attribute int MinorLength(Rmin);
attribute Array<int> Signature[12];
}
    
```

4. 단계적 유사성 측정을 통한 영역 비교

사용자는 객체 모양을 완성한 뒤에, 객체 색상을 시스템의 팔레트로부터 선택한다. 그런 뒤에 사용자 검색 조건인 (1) color-care/don't care, (2) scale-care/don't care (3) shape-care/don't care (4) location-care/don't care를 선택하고 검색 버튼을 누르면 각 특징 값들간의 거리가 순서적으로 계산되고 각각의 가중치의 합에 의해 최종 유사 점수가 계산된다. 이때 4개의 가중치 ($a_1 \sim a_4$)는 1/f로 초기화 된다. 여기서 f는 사용자가 선택한 조건의 수다. 또한 가중치는 사용자에게 의해 조절 될 수 있다.

5. 실험 / 평가 및 향후 연구 계획

FRIP 시스템은 윈도우 환경하에서 Visual C++로 개발 되었으며, off-line 시스템이다. 'query-by-gesture' 검색을 위해서는, 사용자 인터페이스에서 'query-by-gesture' 버튼을 누르고 모니터 옆에 설치된 CCD 카메라에 검색하고자 하는 모양을 그린 뒤에 팔레트로부터 색상을 선택하고 검색 버튼을 누른다. 성능평가를 위해서 본 논문에서는 웹과 Corel-photo CD에서 3000개의 각기 다른 영상들을 이용하였다. 그림 3은 실험을 위하여 임의의 5개 질의 객체를 생성하고(그림 4) 이를 이용하여 검색한 검색율을 보여주고 있다.

그림3에서 보는 것과 같이 태양의 경우는 검색율이 다른 객체들에 비해 우수함을 알 수 있지만, 다른 객체들의 경우 영상 분할 단계 또는 객체들간의 폐색(occlusion)등으로 인해 정확성이 떨어짐을 알 수 있다. 따라서 본 시스템에서는 이러한 단점을 보완하기 위해 향후, 연관성 피드백 학습(relevance feedback learning)을 본 시스템에 적용하여 검색 정확율을 높일 계획이다.

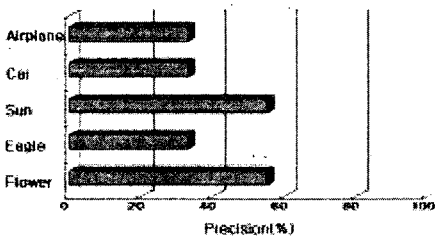


그림3. 5개의 객체에 대한 검색 성능.

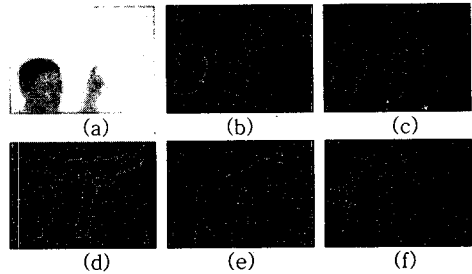


그림 4. 질의 영상 (a) 입력 영상 (b) 해 (c) 꽃 (d) 독수리 (e) 자동차 (f) 비행기

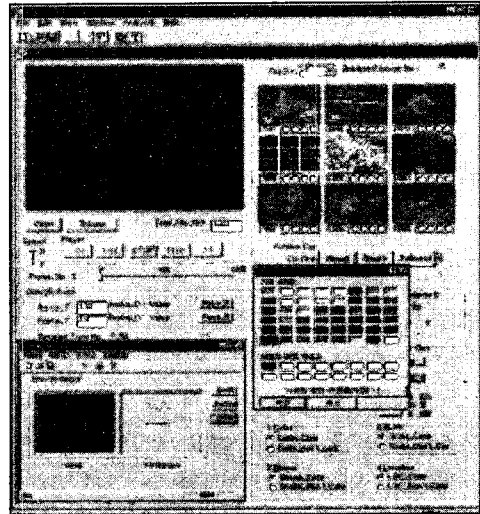


그림 5. 'query-by-gesture'를 위한 FRIP시스템 인터페이스

6. 참고 문헌

- [1] 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.
- [2] Carson, M.Thomas, S. Belongie, J.M. Hellerstein, and J. Malik. "Blobworld: A system for region-based image indexing and retrieval", *In Proc. Int. Conf. Visual Inf. Sys.*, 1999.
- [3] W.Y.Ma, B.S. Manjunath. "Netra: A toolbox for navigating large image database", *IEEE International Conference on Image Processing*, 1997.
- [4] J.R. Smith and S.F. Chang. "VisualSEEK: A Fully Automated Content-Based Image Query System", *ACM Multimedia*, Boston MA, 1996.
- [5] Charles E. Jacobs, Adam Finkelstein, David Salesin, "Fast Multiresolution Image Querying", *Proc. of SIGGRAPH 95*, New York, 1995.
- [6] ByoungChul Ko, Hae-sung Lee, Hyeran Byun, "Region-based Image Retrieval System Using Efficient Feature Description", *Proceedings of the 15th International Conference on Pattern Recognition*, September, Spain, 2000.