

SIFT와 트리구조를 이용한 내용기반 물체인식

주 정 경*, 이 현 창*

Object Recognition using SIFT and Tree Structure

Jung-Kyoung Joo*, Hyun-Chang Lee**

요 약

최근 컴퓨터비전이나 로봇 공학 분야에서 가격이 저렴한 웹캠을 이용한 영상, 즉 2차원 영상으로부터 물체를 인식하는 연구가 활발히 이루어지고 있다. 이러한 로봇이나 비전에서 물체를 찾아내는 여러 가지 방향들이 제시되고 있으며, 지속적으로 로봇은 사람과 유사해져가고 있다. 이를 실현하기 위해서는 사람이 사과를 보고 사과라고 알기 때문에 사과라고 인식하듯이 로봇 또한 미리 알고 있어야 한다는 가정 하에 내용기반의 물체인식이 필요하다. 그러나 엄청난 양의 내용의 데이터베이스가 필요하다. 그래서 용량은 하드웨어기술로 커버가 가능하지만 화면상에 있는 물체들을 빠르게 데이터베이스상의 자료와 매칭이 되어야한다. 본논문에서는 이미지를 SIFT(Scale Invariant Feature Transform)알고리즘으로 BTS(Binary Search Tree)로 트리구조의 데이터베이스를 구축하여 많은 양의 데이터베이스 중 빠르게 검색하여 화면에 있는 물체를 인식하는 방법을 제안하였다.

▶ Keyword : Object recognition, Feature, Search

• 제1저자 : 주정경
* 한세대학교 컴퓨터공학과

I. 서론

최근, 비디오 영상 내의 자동 물체 인식 및 추적기술은 많은 분야에서 응용되고 있으며 국내외에서 관련 기술의 개발이 크게 증가하고 있다. 널리 응용되는 분야는 비디오를 이용한 무인감시 영역으로 보안이 필요한 곳에 감시 카메라의 영상을 입력받아 무단 침입자를 식별하고 지속적으로 추적하게 해준다. 또, 지능형 교통시스템의 구축을 위해서는 특정지역의 교통량을 감시하여 자동차의 통과 대수나 흐름, 사고의 유무 등을 자동으로 수집하여 교통에 대한 정보를 보고해 주는 시스템의 필요성도 크게 증가하고 있다. 이러한 여러 가지 응용의 필요성에 의해 비디오 영상 속에서 사람이나 자동차 등의 움직이는 물체를 자동 검출하고 추적하는 관련 기술은 최근까지 많은 연구가 진행되고 있다.

반면 많은 영상을 검색 하여 물체인식은 진행이 미비하다. 영상을 검색하는 방법은 크게 두 가지가 존재한다. 한 가지는 텍스트 기반 방법이며, 나머지는 내용기반의 검색 방법이다. 텍스트 기반 검색방법은 사용자가 해당 영상과 관련된 적절한 키워드를 추가하거나 검색 시에 관련된 적절한 키워드 추가, 또는 텍스트를 명시적으로 표기 하였다. 텍스트 검색 방법은 초기에 많이 활용되어 졌지만 객관적인 키워드 설정의 어려움, 검색 시 사용자가 키워드를 기억해야한다는 단점 때문에 모호한 경우가 그 한계점이다. 이를 보완 하는 것이 내용기반의 검색방법이다.

내용 기반 검색은 사용자가 검색 질의로 영상 혹은 영상에 대한 정보를 제공하고 이를 분석하여 가장 유사한 영상을 사용자에게 제공하는 방법이다. 이중에 가장 중요한 방법은 질의 정보의 분석이다. 그중 특징 기반 탐색방법의 하나로 SIFT(Scale Invariant Feature Transform) 키포인트를 이용한 물체 인식을 제안하자고 한다.

기존의 템플릿 정합 기반의 물체인식 보다, 변화, 크기, 회전, 노이즈, 조명등 안정성 있게 특징점을 추출한다. 이런 장점을 이용하여 특징점 개수와 영상의 정보를 BST(Binary Search Tree)로 자료구조화시키다. 입력영상이 들어오면 물체를 라벨링하여 특징점 개수를 이용하여 임계값을주어 트리에서 검색한다. 그리고 검색된 자료만 매칭을하여 이 물체가 무엇인지 알려주는 방안을 실험을 통하여 보이도록 한다.

II. 특징점 추출과 매칭, 이진탐색트리

2.1. 특징점 추출

미래의 로봇은 물체 인식 등 동적인 환경에서 강인한 특징점을 검출해야한다. 이런 불변한 특징을 검출하는 방법은 SIFT(Scale Invariant Feature Transform) 알고리즘이다. 이 알고리즘을 이용하여 영상 변화, 크기, 회전, 노이즈 영향에도 일정한 특징점 정합 성능을 유지하며 특징점 추출 방법은 크게 세 가지 단계로 구성되어 있다.

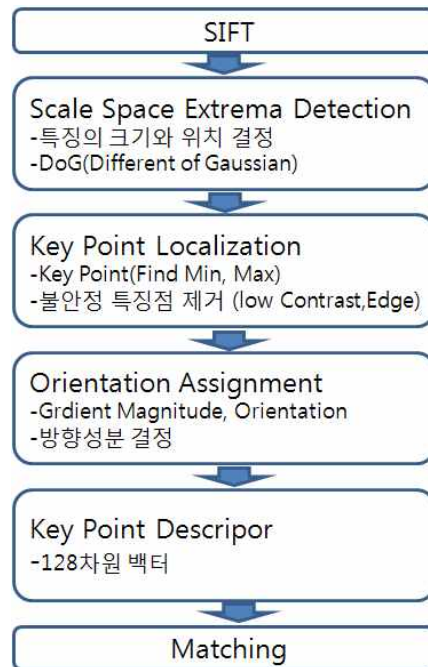


그림 1. SIFT 구성도

첫 번째 단계는 스케일 공간상의 후보 특징점을 추출하고 후보 특징점의 안정성을 검사하여 안정된 특징점의 위치를 세 부 위치로 보정한다. 두 번째 단계는 특징점에 기준 방향을 할당한다. 세 번째 단계는 기준 방향을 중심으로 특징점 주변 영역의 국부 영상에서 특징점 기술자를 생성한다. 네 번째 단계는 정규화 과정을 거치며 이렇게 만들어진 128차원 벡터는 특징점을 대표하는 기술자가된다.

후보 특징점을 추출하기 위해 모델 영상의 스케일 공간을

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \dots\dots\dots (1)$$

와 같이 정의한다. (1)에서 L은 가우시안 블러링 영상을

의미하며, I는 입력영상 모델이미지이며, G는 가우시안 함수의 컨볼루션으로 표시한다.

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \dots\dots\dots (2)$$

와 같이 정의한다. 가우시안 차분 방법은 영상과 컨볼루션된 가우시안 함수들의 차분에서의 극점들을 이용하여 특징점들의 안정적인 위치를 효율적으로 검출하는 방법이다.

가우시안 차분식은 스케일 공간상에서 안정된 특징점을 찾기 위해서는 DoG

(Difference of Gaussian) 함수 D를

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \dots\dots (3)$$

$$= L(x,y,k\sigma) - L(x,y,\sigma)$$

와 같이 정의한다. 여기서 k는 보통 $\sqrt{2}$ 를 사용하여 k배 만큼 점진적으로 크게하여 입력 영상의 블러링 정도를 점진적으로 증가시킨다. 그림 1과 같이 블러링 영상들의 이웃 영상들을 이용하여 차 영상(DoG)을 구하고 각각의 차 영상의 모든 화소들을 그림 2와 같이 비교한다.

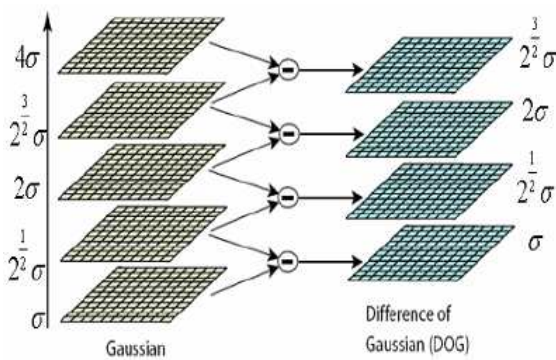


그림 2 스케일 공간상의 DoG영상

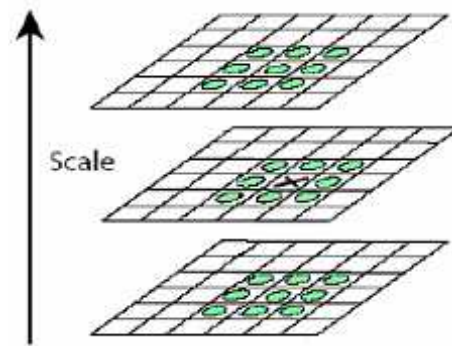


그림 3 Dog 영상 스케일 공간상의 후보 특징점

그림 2는 Dog 영상 내에서 후보 특징점을 찾는 과정을 나타낸다. 현재위치의 화소를 중심으로 이웃한 화소와 이웃 DoG 영상에서의 같은 위치 화소 주변의 3×3 영역에 포함되는 26개 화소의 값을 현재 화소와 비교하여 최소의 값 또는 최대의 값 이면 후보 특징점으로 등록한다. 이 방법으로 찾아낸 후보 특징점들은 주변 화소들에 따라 안정성이 정해진다. 일단 안정적인 특징점으로 정해지면 후보 특징점은 세부 위치로 이동하게 된다.

두 번째 단계에서는 가우시안 영상에서의 주변 픽셀의 차이를 이용하여 기준방향(Orientation)과 크기를 가진 벡터를 생성한다. 특징점의 스케일 성분을 이용하여 스케일 공간상의 블러링 영상을 선택하고 특징점 위치의 주변 상하좌우 화소를 이용하여 방향과 크기를 가진 벡터를

와 같이 정의하고 식(4)에서 $m(x,y)$ 와 $\theta(x,y)$ 는 주변 화소를 이용하여 특징점에 크기와 방향을 가진 벡터를 나타낸다. 이렇게 만들어진 특징점 벡터는 영상의 회전이 있더라도 특징점 주변 화소의 값을 변하지 않기 때문에 영상회전에 대해서 일정한 방향성질을 유지할 수 있다. 영상이 회전한 만큼 특징점의 방향도 같이 회전하게 되는 특징점 기준 방향 성질은 영상 회전에 변하지 않는 특징점 기술자를 만들기 위해 중요한 기준이 된다.

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \dots\dots (4)$$

$$\theta(x,y) = \tan^{-1} \left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)} \right)$$

특징점 기술자는 모델 영상과 물체와의 특징점 정합에 기준이 되는 성분이다. 영상의 크기, 회전, 노이즈와 같은 외부 환경에 의한 영상 변화에도 강인한 특징점 정합 성능을 유지한다. 우선 특징점 주변영역에서 특징점의 방향을 할당하는

방법과 같이 특징점 주변의 국부 영역의 그라디언트의 방향과 크기를 구한다. 영상회전에 불변하는 기술자를 얻기 위해 그라디언트의 방향과 기술자의 좌표는 특징점 방향을 기준으로 회전이동이 이루어진다.

이렇게 회전된 그라디언트 영상에 특징점을 중심으로 한 가우시안 가중치를 부여한다. 가우시안 가중치의 목적은 특징점 주변 영역의 작은 변화에도 그라디언트의 방향과 크기가 민감하게 변하는 것을 피하기 위함이다. 특징점을 중심으로 가우시안 가중치를 부여함으로써 특징점 주변 영역의 그라디언트를 강조하여 에델르 최소화한다. 그림 3에서 볼 수 있듯이 가우시안 가중치가 부여된 이미지 그라디언트는 각각의 배열이 8방향을 가지고 4×4배열로 재구성된다. 이러한 방향성 히스토그램은 128(8×4×4)차원의 벡터로 구성되며 조명에 강인한 기술자를 만들기 위해 정규화 과정을 거치며 이렇게 만들어진 128차 원 벡터는 특징점을 대표하는 기술자가 된다. 또한, 회전에 강인하게 하기 위해 기존의 사각 기술자를 원형으로 생성하였다.

2.2. 특징점 매칭

DoG는 모서리 및 그 주변에 강하게 반응하며, 이 부근에서 추출된 서로 다른 물체의 특징점들의 특성 해석 결과는 비슷할 수 있다. 이는 정합 절차에서 시스템의 의사 결정을 어렵게 한다. 또한, 모델 물체의 특징점을 추출하기 위해 Dog의 극한값을 찾을때, 배경 부분과 대비가 작은 영역에 특징점들이 추출되는 것을 볼 수 있다. 이렇게 추출된 특징점들은 배경과 대상 모델의 구분을 어렵게 할 뿐 아니라, 의미 있는 정보를 제공하지 못한다. 또한 다음 인식 단계에서 불필요하게 많은 점들의 정합은 인식 시스템의 속도와 인식율을 저하시킨다. 이것은 Hessian행렬의 곡(curvature)을 사용한 방법으로 모서리 영향을 감소시킬 수 있다. 입력 모델 영상에서 SIFT로 검출한 특징점은 실사 영상에서 같은 방법으로 검출한 특징점과 정합을 시도한다. 모델 영상과 물체가 가진 SIFT 특징점 기술자를 이용한 특징점 정합은

$$d(A,B) = \sqrt{A^2 - B^2} \dots\dots\dots (5)$$

와 같이 정의 하며, 식 (5)는 유클리드 거리를 이용하여 SIFT의 특징점 사이의 거리를 구하는 식이다. 일반적으로 유클리드 거리를 이용한 특징점 정합 방법에 있어서는 여러 데이터 중에서 유클리드 거리가 가장 가까운 데이터를 정합으로 선택하는 경우가 발생한다. 그러나 단순히 거리가 가장 가까

운 점으로 정합된 점으로 판단할 경우 비슷한 거리의 특징점의 2개 이상일 경우에는 정합 데이터 중에서 오정합된 특징점이 발생할 확률이 높다. 이러한 오정합의 비율을 줄이기 위해

$$\frac{d(f^A, f^C)}{d(f^A, f^{C1})} \leq 0.5 \dots\dots\dots (6)$$

과 같이 유클리드 거리의 비율을 사용한다. 여기서 f^A 는 비교할 기준 특징점의 위치를 나타내며, f^C 는 정합 중에 가장 좋은 정합 결과를 갖는 특징점의 위치를 나타내며, 그리고 분모에 f^{C1} 는 두 번째로 좋은 정합의 결과를 갖는 특징점의 위치를 나타낸다.

2.3. 이진탐색트리

이진트리 기반의 탐색을 위한 자료구조이며 일상생활에서도 전화번호를 찾거나 사전을 찾거나 많은 부분에 응용이 된다.

이진 탐색 트리란 모든 원소의 키는 유일하며, 왼쪽 서브 트리의 키들은 루트의 키보다 작다. 오른쪽 서브 트리의 키들은 루트의 키보다 크며, 왼쪽과 오른쪽 서브 트리도 이진 탐색 트리이다.

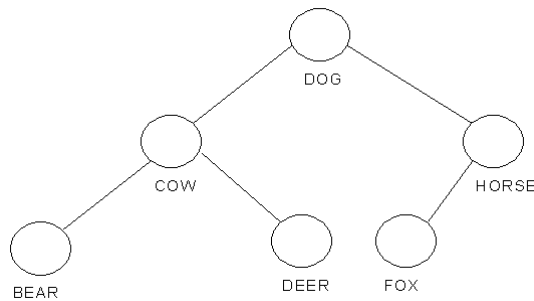


그림 4. 이진트리

탐색할때 루트Ni, 탐색키 K라면 공백트리일때 탐색 실패, 종료하고 k와Ni가 같다면 Ni가 목표 노드가 되며, k 가 Ni보다 작을때 Ni <- ROOT(LT(Ni)) 되고반대로 Ni보다 클때는 Ni <- ROOT(RT(Ni))로 탐색한다.

삽입은 리프노드의 위치의 삽입이 일어나며 루트가 Ni, 삽입 레코드의 키가 K 라고 하였을때 공백 트는 K의 노드가 루트가 되며 K 와 Ki가 같을때는 삽입 실패를 하게된다. K 가 Ki 보다 작을 때 는 왼쪽 서브트리 탐색하게 되며 Ki보다 클때는 오른쪽 서브트리 탐색한다.

삭제는 리프의 삭제시 노드 제거를 하며 리프가 아닌 경우

서브트리 유지하고 삭제 표시

성능 자주 접근되는 노드를 루트에 근접시켜 균형트리로 만들며 트리의 최대 경로 길이를 최소화하며 N노드는 $\lfloor \log_2 n \rfloor + 1$ 이다.[4]

III. 트리 구성 및 물체 인식

3.1. 트리구성

본 논문은 입력영상에 대해 빨리 입력영상에 대해 라벨링하여 비교하는 것도 중요하지만 그보다 중점을 둔 사항은 미리 구성된 트리구조의 데이터베이스에서 빠르게 물체를 찾는 것이다. 그래서 트리구성이 중요한데 트리를 구성하는 정보는 이미지를 한물체에 대하여 적게는 두 가지 방향에서 바라본 영상이 필요하며 또한 경우에 따라 필요하다면 5방향에서 바라본 영상도 만들어 소스이미지를 만든다. 이렇게 준비되어진 소스 이미지를 특징점의 개수 및 정보를 SIFT를 이용하여 추출되면, 그다음의 특징점의 개수를 기반으로 이진탐색트리 구조로 구축한다. 이렇게 구축되어진 트리구조의 데이터베이스를 가지고 프로그램은 물체인식에 속도를 높이게 된다.

3.2. 물체 인식

입력영상을 전처리 과정(그레이로 변환, 가우시안 블러링, 모폴로지, Canny경계선 검출)을 거쳐 라벨링하여 라벨에 따라 물체를 하나 하나의 다른 이미지로 만들게 되며 각 이미지를 SIFT를 이용하여 특징점의 개수를 추출해 낸다. 그렇게 나온 특징점 개수에 대한 임계값을 정해 트리구조에 있는 특징점 개수를 탐색하게 된다. 탐색되면 그 노드에 자료를 이용하여 그 이미지와 입력이미지를 매칭을 하여 특정 이상의 매칭 개수가 되면 물체를 물체라고 인식하는데 SIFT를 이용한 영상 추적 알고리즘[6]논문에서는 3개 이상이 적당하다고 실험 하였으나 실험 도중 특징점이 적게 나오는 물체에 대해서는 매칭되는 개수도 적게 될 수 있었기 때문에 본 논문에서는 소스이미지의 특징점 개수의 10% 이상이 매칭 되면 물체를 인식하게 된다. 인식된 물체는 물체의 고유에 이름을 화면상에 출력하게된다.

IV. 실험 및 결과

본 연구는 내용기반 물체 인식 방법을 사용하고 있다. 데이터베이스를 구성하기 위해 물체의 정면영상과 물체에 따라

위영상 옆영상의 특징점개수와 부가적 정보를 추가적으로 데이터베이스에 저장한다. 이 후 입력영상이 주어지면 입력된 영상을 라벨링하여 물체의 특징점의 개수를 가지고 트리구조의 데이터베이스에서 임계값범위에 따라 검사한다. 그중에 검색이 되면 검색된 영상과 정합을 한다. 여기서 물체의 특징점의 매칭은 10% 이상이어야 정합이라 판단하였다. 사용한 시스템은 영상처리 라이브러리 OPENCV와 수치라이브러리 GSL을 사용하여 Visual C++을 사용하여 이 논문에서 제안한 알고리즘을 구현하였다. 입력 영상의 크기는 320×240 이며, 각 트리구조의 데이터베이스에 저장되는 이미지는 각각 크기가 다르다.

V. 결론 및 추후과제

기존의 방법은 하나의 이미지 영상을 보여주고 그 영상 안에는 하나의 물체를 추적하거나 하나를 찾는데 중점을 두었으나 물체의 회전, 물체의 크기 변화 등에 대하여 강인하게 수행할 수 있다는 장점이므로 트리구조의 데이터베이스기반으로 빠른 검색을 시도하여 물체를 인식 하였다. 앞으로 로봇에 들어오는 이차 영상만으로 물체를 판단하는 등 응용 할 수 있다. 그러나 SIFT 알고리즘의 계산 성능계선을 하여 실시간으로 구현해야할 과제이다.

참고문헌

- [1] RC. Gonzalez, RE. Woods, "Digital Image Processing", Addison Wesley, pp.433-455, (1992)
- [2] D.H.Ballard, Computer Vision, Prentice-Hall, Inc., pp.76-79, 1991.
- [3] McAndrew. Introduction to image processing with MATLAB, Thomson
- [4] David G. Low, "Distinctive Image Features from Scale-Invariant KeyPoint", International Journal of Computer Vision, Vol 20 pp. 91~110, 2003
- [5] A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing," vol. 13, no. 4, pp. 471-494, October 1966.
- [6] 천인국, Data Structures in C, 생능출판, 2005
- [5] 황선규, IT EXPERT 영상 처리 프로그래밍 by Visual C++, 한빛미디어, 2007

- [7] 한양대 대학원, SIFT를 이용한 영상 추적 알고리즘
2007 TM 621.367 7544s v, 38 p.
- [8] 강석천, SIFT를 이용한 위성사진의 정합 기법, 중앙대
대학원 2007 TM 006.42 7257s v, 57 p.
- [9] OpenCV KOREA, <http://cafe.naver.com/opencv.cafe>