

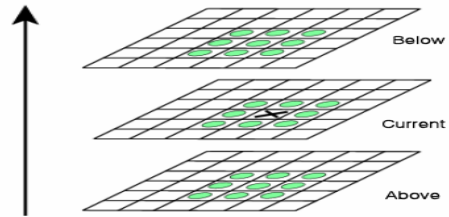
SIFT알고리즘을 이용한 물체인식

윤준영\*, 김은태\*\*, 전세웅\*\*\*  
연세대\*, 연세대\*\*, 전자부품연구원\*\*\*

Object recognition using SIFT algorithm

Joon-Young Yun\*, Eun-Tae Kim\*\*, Se-Woong Jeon\*\*\*  
Yonsei University\*, Yonsei University\*\*, Korea Electronics Technology Institute\*\*\*

**Abstract** - 본 논문은 Scale Invariant Feature Transform(SIFT)알고리즘으로부터 얻어진 로컬 특징점으로부터 물체를 인식하는 방법에 대하여 논하였다. SIFT알고리즘은 물체의 스케일, 회전에 강인하고, 또한 3차원 시점의 변화에도 부분적으로 강인한 특징점을 추출한다. SIFT 알고리즘은 입력영상에 크기가 다른 가우시안 함수를 적용하고, 블러링된 영상들의 차 영상에서 극값을 추출하여 특징점으로 사용한다. 하지만 SIFT알고리즘에서 가우시안 함수를 적용하는 것은 상당히 많은 연산을 필요로 하기 때문에 본 논문에서는 하나의 옥타브를 사용하여 연산시간을 단축하였다. 하나의 옥타브를 사용함으로써 물체의 스케일이 크게 변하였을 때는 문제가 발생한다. 이를 해결하기 위하여 대상 물체의 작은 스케일, 큰 스케일에서 추출된 특징점을 혼합하여 DB를 생성하였다.



<그림 1> DoG를 이용한 지역적인 극값 검출 방법 (x표시를 한 화소가 O표시를 한 주위의26개의 화소보다 크거나 작으면 극값으로 선택한다.)

1. 서 론

물체인식은 컴퓨터 비전 시스템에서 풀어야할 가장 어려운 문제중의 하나이다. 정형화 되어있지 않은 환경에서 물체인식 시스템을 구현하려면, 물체가 또 다른 물체로부터 가려지는 현상(occlusion), 물체의 크기가 변하는 현상, 물체가 회전하여 x, y좌표가 바뀌는 현상 그리고 보는 시각에 따라 달라지는 물체의 변화(어파인 변환)에 대하여 목표 물체를 식별할 수 있어야 한다. 물체인식에 대한 기존의 방법들, 즉 에지와 같은 글로벌한 정보에 의존하는 방법은 occlusion과 같은 현상에 매우 취약함을 보인다. 반면에 지역적인 특징을 이용하여 물체를 인식하는 방법은 로컬 특징점 주위의 지역적인 공간에만 영향을 받기 때문에 occlusion 현상이 발생하여도 로컬 특징점을 추출하는데 아무런 영향을 주지 못한다. 이런 로컬 특징점은 1) 여러 스케일에서 가우시안 미분 함수를 사용하여 코너, 원형, 접합점등의 특징을 추출하는 방법[3, 4, 5], 2) 가우시안 함수가 적용된 영상들의 차에서 로컬 극값을 특징점으로 추출하는 방법[6]으로 크게 구분된다. 가장 인기있는 로컬 특징점 추출 방법은 1988년에 Harris에 의하여 제안된 해리스 코너 디텍터[2]이다. 해리스 코너 디텍터는 물체의 회전변화에는 매우 강인함을 보인다. 하지만 스케일의 변화에는 강인하지 못한 단점이 있다. 이는 2004년 D.W. Lowe에 의해 제안된 Scale Invariant Feature Transform(SIFT)[1]에 의하여 해결되었다. SIFT알고리즘은 총 4단계로 나뉘어 구현된다. 1단계(Scale space extrema detection)에서는 서로 다른 스케일에 가우시안 함수를 적용하고, 차 영상에서 극값을 추출하여 후보키로 선정한다. 2단계(Keypoint localization)에서는 1단계에서 잘못 추출된 후보키들을 제거한다. 3단계(Orientation assignment)에서는 추출된 키포인트들에 방위와 크기를 할당한다. 4단계(Keypoint descriptor)에서는 키포인트들에 128차원의 벡터를 생성한다. 우리는 이상의 단계를 거쳐 추출된 특징점을 이용하여 물체인식에 사용하였다.

2. 본 론

2.1 Scale Invariant Feature Transform(SIFT)

SIFT는 먼저 서로 다른 크기의 분산 값을 갖는 가우시안 함수를 적용한 영상  $L(x, y, \sigma)$ 들의 차이를 이용해 얻어진 가우시안 차분영상  $D(x, y, \sigma)$  즉, DoG(Difference of Gaussian)를 통해 <그림 1>에서 보는 바와 같이 로컬 극값(후보키)을 구한다.

$$L(x, y, \sigma) = G(x, y, \sigma) \times f(x, y) \quad (1)$$

$$G(x, y, \sigma) = 1 / (2\pi\sigma^2) \times e^{- (x^2 + y^2) / 2\sigma^2} \quad (2)$$

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3)$$

이렇게 구해진 후보키들은 키포인트 정교화(Keypoint localization)과정을 거쳐 안정하다고 판단이 되면 키포인트로 선택된다. 키포인트 정교화 과정은 다음과 같다.

식(4)는 원점이 해당 후보키의 위치에 놓이도록 이동된  $D(x, y, \sigma)$ 에 대한 테일러급수를 나타낸다. 여기서  $D$ 와  $D$ 의 도함수는 후보키의 위치에서 평가되며  $X = (x, y, \sigma)^T$ 는 현재 후보키의 위치로부터 오프셋을 의미한다. 만일 오프셋이 0.5보다 크다면 극값은 이웃화소에 근접해 있음을 의미한다. 실제 극값의 위치  $\hat{X}$ 는  $X$ 에 대한 도함수를 구하고 그 값을 0으로 하여, 식(5)와 같이 정의하여 구할 수 있다. 식(5)의 피크에서의  $D(\hat{X})$ 는 낮은 contrast를 갖는 불안정한 피크를 제거하는데 사용된다. 키포인트의 안정성을 보장하기 위해서는 앞에서 기술한 낮은 contrast를 갖는 후보키들을 제거하는 것만으로는 충분하지 않다. DoG는 실제 엣지에 대해 큰 값을 갖는다. 하지만 후보키중에서 엣지를 따라서 DoG값이 적은 값을 갖는 것들은 적은 양의 노이즈에도 민감하게 반응하기 때문에 불안정한 점으로 판단하고 제거되어야 한다. 이는  $2 \times 2$  Hessian 행렬의 고유치로써 판단할 수 있으니 Harris, Stephens이 사용한 방법을 이용하면 고유치를 구할 필요 없이 대각합과 행렬식을 이용하여 판단할 수 있다.

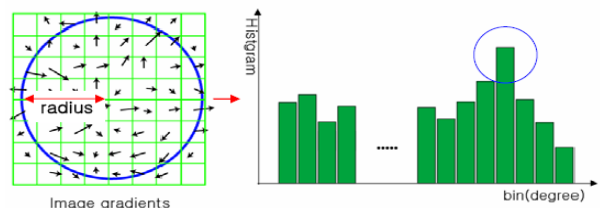
$$D(X) = D + \frac{\partial D^T}{\partial X} + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \quad (4)$$

$$\hat{X} = - \left( \frac{\partial D}{\partial X^2} \right)^{-1} \frac{\partial D}{\partial X} \quad (5)$$

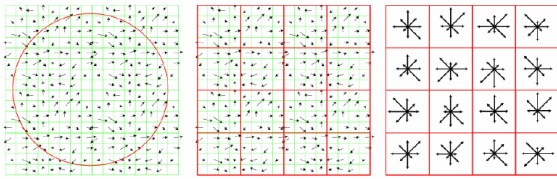
키포인트(특징점)가 선택되었으나 다른 영상에서 얻어진 특징점과 비교를 하기 위해서 서술자(descriptor)가 필요하다. 다음의 식 (6), (7)을 이용하여 특징점주위의 화소들의 크기와 방향을 구한 후에 <그림 2>와 같이 주위화소들의 방향을 36개의 계급(bin)으로 나누고 방향과 크기를 히스토그램에 추가한다. 히스토그램의 가장 큰 값을 갖는 방향을 특징점의 주 방향으로 정한다. 구해진 주 방향과 크기는 서술자를 얻는 과정에 사용된다.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (6)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, Y) - L(x-1, y))) \quad (7)$$



**<그림 2> 오리엔테이션 히스토그램 구성 ( $\theta$ 에 크기를 곱하고 36bin에 대하여 양자화 한 후에 영역에 대하여 히스토그램을 구성한다. 히스토그램의 최대값이 특징점의 주 방향이된다.)**



**<그림 3> 128차원의 서술자 생성**

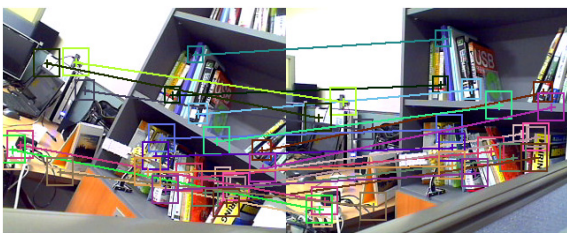
<그림 3>의 왼쪽과 같이 특징점이 추출된 DoG영상의  $\sigma$ 의 1/2에 해당 하는 반지름을 갖는 가우시안 분포를 가중치로 하여  $16 \times 16$ 의 화소들을 8방향의 히스토그램에 누적하여  $4 \times 4$ 의 서술자를 구한다. 이때 영상이 회전하여도 영향을 미치지 않게 하기 위하여  $16 \times 16$ 의 주변 화소 값들은 특징점의 주방향에 맞춰 회전시킨다. 이로써  $4 \times 4 \times 8 = 128$ 차원의 서술자를 얻게된다. 서술자들은 유클리디안 거리를 비교하여 매칭된다.

**2.2 SIFT알고리즘의 성능 향상**

SIFT는 연산량이 많은 가우시안 함수를 사용한다. 가우시안 함수는 영상의 크기가 크면 클수록, 분산의 크기가 크면 클수록 연산량은 기하급수적으로 늘어나게 된다. 그리하여 가우시안 함수를 1차원 커널로 적용하는 방법도 있지만 이 또한 영상의 크기가 커지면 마찬가지로 느려질 수 밖에 없다. 제안된 방법은 옥타브를 제한하여 연산량을 줄여 속도를 빠르게 하였다. 옥타브를 제한하여 SIFT알고리즘을 사용할 경우 대상 물체 혹은 장면의 스케일의 변화가 클 경우에는 매칭 성공률이 매우 떨어지는 현상이 발생한다. 이는 특징점의 서술자를 생성할 때 스케일 변화의 정보를 서술자에 반영시키지 못하고 현재 스케일의 정보만 반영시키기 때문이다. 이러한 문제점을 해결하기 위하여 대상 물체의 서술자 DB를 생성할 때, 물체의 큰 스케일에서의 서술자와 작은 스케일에서의 서술자 따로 생성하고, 생성된 서술자들을 혼합하여 DB에 저장한다. 이렇게 함으로써 DB는 한 물체에 대한 여러 스케일의 정보를 갖는 서술자를 얻게된다.

**2.3 실험**

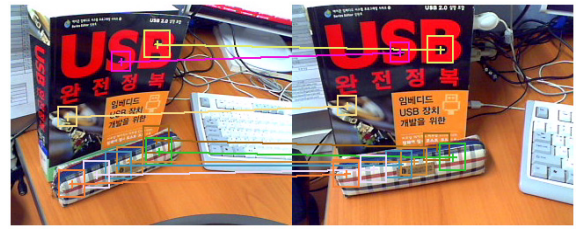
제안된 기법은 Core2Quad Q6600, 2G 메모리의 Desk-top 환경에서 구현하였다. 연구실 주위 환경 인식과 스케일 변화가 있는 물체(필통) 인식을 실험하였다. <그림 4>는 회전, 어파인, 작은 스케일변화에 따른 주위 환경의 매칭 결과를 보여준다. 그림에서 보는바와 같이 회전 변화, 작은 스케일 변화에는 매우 강한 모습을 보였고, 어파인 변화에는 매칭 되는 특징점이 줄어들기는 하였으나 상당한 매칭성공률을 보였다. <그림 5>는 물체인식 실험의 결과이다. (a)와 같이 스케일의 변화가 심하게 일어났을 때에는 옥타브를 제한하여 스케일 변화에 대처를 하지 못하였지만 (b)는 제안한 알고리즘을 적용한 결과이다. 다양한 스케일정보를 DB가 가지고 있기 때문에 스케일의 변화에 잘 대응한 것을 볼 수 있다.



(a)

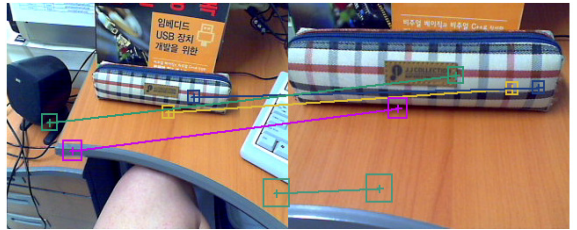


(b)

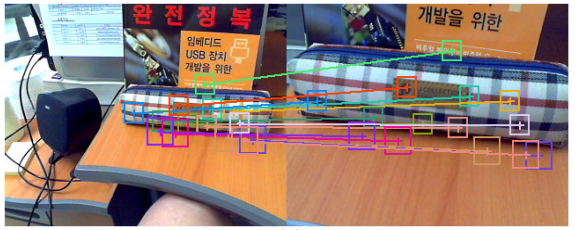


(c)

**<그림 4> (a) 회전 변화에 의한 매칭 결과, (b) 스케일 변화에 의한 매칭 결과, (c) 어파인 변화에 의한 매칭 결과**



(a)



(b)

**<그림 5> 큰 스케일의 변화가 있을 때의 매칭 결과 (a) 다양한 스케일의 서술자를 혼합하여 DB를 생성하지 않았을 경우, (b) 다양한 스케일의 서술자를 혼합하여 DB를 생성하였을 경우**

**3. 결 론**

본 연구에서는 [1]에서 제안된 Scale Invariant Feature Transform(SIFT)알고리즘의 성능에 많은 영향을 주는 큰 영상에 가우시안 함수를 적용 시키는 것을 제한함으로써 좀 더 빠른 SIFT알고리즘을 구현하였다. 하지만 성능을 향상시키기 위해 옥타브를 제한함으로써 한 장면을 매칭 하는데 있어서 스케일의 변화에 매우 민감한 반응을 보였다. 하지만 제안한 DB혼합 기법을 사용함으로써 큰 스케일 변화에도 매우 강한함을 보였다. 현재 비전 시스템에서 회전, 스케일의 변화에 강인함을 보이는 특징점 추출 방법은 상당한 수준이다. 하지만 실제 인식하는 과정에서 회전, 스케일의 변화뿐만 아니라 어파인 변화도 큰 비중을 차지한다. 어파인 변화에 불변하는 특징점 추출 방법은 매우 어렵다. 앞으로는 어파인 변화에 불변하는 특징점 검출 방법에 대한 연구가 비전 시스템에서 큰 관심사가 될 것이다.

**[참 고 문 헌]**

[1] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", Int. Journal of computer Vision, 60(2), 91-110, 2004  
 [2] C. Harris and M. J. Stephens, "A combined corner and edge detector", In Alvey Vision Conference, 147-152, 1988  
 [3] T. Lindeberg, "Feature Detection with Automatic Scale Selection", Int J. of Computer Vision, Vol.30, No2, pp.77-116, 1998  
 [4] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points", Proc. of the 8th Int Conf. on Computer Vision, pp. 525-531, 2001  
 [5] Y. Dufournaud, C. Schmid, and R. Hor명, "Matching Images with Different Resolutions", Proc. of the Conf. on Computer Vision and Pattern Recognition, 612-618, 2000  
 [6] D. Lowe, "Object Recognition from Local Scale Invariant Features", Proc. of the Int. Conf. on Computer Vision, pp. 1150-1157, 1999