

1. Reviewer : 하야로비
2. Article : David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

3. Overview :

A. About Scale Space

- i. 우리가 어떤 물체의 크기(특징)에 대해 이야기 할 때 적합한 척도에서 물체의 특징을 설명함. 예를 들어, 나뭇가지를 설명하는데 크기가 몇 센티미터 또는 몇 미터이다라고 이야기함. 즉 물체마다 물체의 특징을 설명하기 위한 적합한 척도(크기, scale)가 있음.
- ii. 현실의 물체는 다양한 척도를 가질 수 있음. 적합한 척도를 알 수 있음. 하지만 영상에서 적합한 척도(scale)를 알기 어렵기 때문에 다양한 척도(multi scale)로 물체를 표현 → Scale Space라고 함
- iii. 영상(digital image)을 Scale Space에서 표현하기 위하여, Gaussian Kernel적용 ← 유일한 Kernel
- iv. 결국, 영상의 척도(Scale)에 강건한 특징을 추출하기 위해 Scale Space 사용
- v. Critique
- vi. Questions and Issues

B. Generation Scale Space

- i. 척도에 불변한 특징을 추출하기 위해 척도 공간에서 극점(극대점 또는 극소점)을 찾아야 함. 이를 위해 Laplacian 을 이용 (Laplacian에 가장 근사화 된 DoG 사용, 처리속도가 빠름)
- ii. Scale Space 생성을 위해 원본 영상에 Gaussian 적용. 한 옥타브 내에서 initial image와 final image의 Gaussian의 크기가 2배가 되도록 Gaussian의 크기를 설정. 한 옥타브에 s의 구간을 갖도록 영상을 생성
- iii. 인접한 Gaussian Blurred image의 차로 DoG 생성
- iv. Critique
- v. Questions and Issues
 1. DoG의 극점이 영상의 에지 부분인가요? 아니면 흐려지는 일부 영역의 중심점 인가요? 만약 중심점이라면 왜 중심점에서 Keypoint를 찾는 것일까요? 그렇게 찾아진 중심점은 Scale에 불변한 중심점인가요? 즉 Scale이 변한다고 해도 항상 같은 위치에서 찾아질까요?

C. Scale Space Extrema Detection

- i. 적합한 Scale을 찾기 위해, 한 옥타브내에서 현재 3x3주변 픽셀과 이전 Scale•다음 Scale의 3x3 주변 픽셀의 밝기 값을 비교해서 현재 픽셀의 밝기 값이 가장 큰 값이거나 가장 작은 값인 경우 Keypoint로 결정
- ii. 이 단계에서 선택된 Keypoint의 Scale과 가장 유사한 Scale에서 Keypoint Descriptor를 생성
- iii. Critique
 1. Scale을 3x3을 선택한 것이 최선일까요? 이 부분을 Adaptive하게 할 수는 없을까요?(논문에서도 이부분은 적당한 커버 영역이 없다고 언급하고 있기는 합니다만 정말 방법이 없는 걸까요?) Keypoint를 선택하기 위해 현재 Scale과 전,후 Scale을 사용하고 있고 그 이유는 실험적인 결과로 그렇게 선택을 했다고 했는데 노이즈가 많은 영상의 경우에도 이게 적용 가능할까요? 논

문에서는 Keypoint Selection이 영상의 종류에는 크게 영향이 없다고 언급하고 있지만 초음파 영상과 같이 노이즈가 심한 경우에도 정말 Keypoint Selection이 상관 없을가요?

iv. Questions and Issues

D. Keypoint Localization

i. C.i.에서 선택된 Keypoint의 정확성을 높이기 위해, Location, Scale, Edgeness정도에 따라 Keypoint를 선택

ii. Location Correlation

1. 논문 4절의 수식 2번에서 결국 x 가 Location과 Scale에 대한 벡터이므로 $D(x)$ 의 미분 값이 0이 되는 위치가 Keypoint의 Location이 되고, 만약 미분값에 대한 Offset이 0.5보다 큰 경우에는 x 의 위치가 다른 sample point의 위치와 더 가까운 것이므로, 다른 sample point를 interpolation함.

iii. ii.1.에서 찾은 위치에 있는 Keypoint에 대해 Contrast가 0.03보다 작을 경우(밝기 값을 0-1로 정규화) Keypoint 후보군에서 제거

iv. 노이즈가 있는 부분이나 약한 에지 등에서 찾아진 Keypoint제거, 이를 위해 Principal Curvature가 일정한 크기 이하가 되는 경우는 Keypoint 후보군에서 제거, Principal Curvature를 구하기 위해 Hessian Matrix 이용, Principal Curvature가 H Matrix의 고유값과 비례한다고 함. 따라서 고유값의 비율이 일정 이상을 넘지 않는 keypoint만 남김. 논문에서는 두 고유값의 비율이 10임.

v. Critique

vi. Questions and Issues

E. Orientation Assignment

i. C, D.에서 구한 Keypoint 후보군의 방위를 결정하기 위해, Keypoint의 Scale과 가장 가까운 Scale을 가지는 Gaussian Blurred Image에서 픽셀의 기울기와 방위를 구함. 이때 GBI의 모든 픽셀에 대해 기울기와 방위를 구하고, GBI의 Scale(결국 Original Image에 얼마나 큰 sigma를 적용했는가를 알아내는 거겠조?)에 1.5배가 되는 sigma를 가지는 가우스 가중치 윈도우를 사용해서 기울기의 크기를 더함. 결국 36개의 히스토그램 중에서 픽크가 되는 방위를 Keypoint의 방위로 사용. 이때 피크의 크기에 80%내에 포함되는 히스토그램의 값을 같은 위치에서 다른 방위에 대한 Keypoint로 사용.(결국 같은 위치에서 다른 방위를 가지는 Keypoint가 생성 될 수 있음.)

ii. Critique

iii. Questions and Issues

F. Keypoint Descriptor

i. Keypoint를 중심으로 16x16 범위 내에서 Keypoin Descriptor를 생성함. 기본적으로 유의할 점은 Keypoint선택은 DoG에서 이루어지고 실제 Keypoint Descriptor는 Keypoint와 유사한 Scale를 갖는 Gaussian Blurred Image에서 수행됨.

ii. 회전 불변 특징을 생성하기 위해 GBI영상의 모든 픽셀의 방위를 Keypoint좌표로 이동(GBI 영상의 기울기 - Keypoint의 방위)

iii. Keypoint주변의 픽셀의 기울기 방위를 8개 방향의 히스토그램으로 변환 한 값이 특징점(Keypoint

Descriptor)이 됨

iv. Critique

v. Questions and Issues

1. Subregion의 Histogram을 계산 할 때 Boundary Effect라는게 정확히 뭔지 이해가 안됩니다.

그 문제를 없애기 위해서 trilinear interpolation을 한다고 하는데, 왜 그런건가요?

G. Matching에 대한 부분은 생략하였습니다. 이 부분은 여러 가지로 응용될 수 있는 부분들이기 때문에 여러 가지 정합 방법이 있을 수 있다고 생각합니다. 다만 그 흐름만 정리해 보자면,

i. Reference Keypoint(인식하고자 하는 물체의 Keypoint를 미리 생성해서 DB 구축)와 Target Keypoint의 거리가 최소가 되는 Keypoint의 풀을 만듦. 이때 단순히 거리가 최소가 되는 Keypoint를 사용하지 않고 최소가 되는 Keypoint와 그 다음으로 최소가 되는 Keypoint의 거리의 비율을 이용함. 이 때 비율이 0.8이상인 Keypoint는 제거함.

ii. BBF 알고리즘을 사용해서 각 Keypoint를 인덱스 함.

1. 정확하게 이해 못했음. i.에서 찾아진 keypoint list에 대해 정렬하는 걸로 이해하고 있음.

iii. 일차 선별된 Keypoint들에 대해 Generalized Hough Transform을 이용해서 3개의 클러스터를 구성, Affine Transform만을 고려하고 있는 것으로 보임(Affine Transform을 찾기 위해 3개의 점이 필요)

4. Critique :

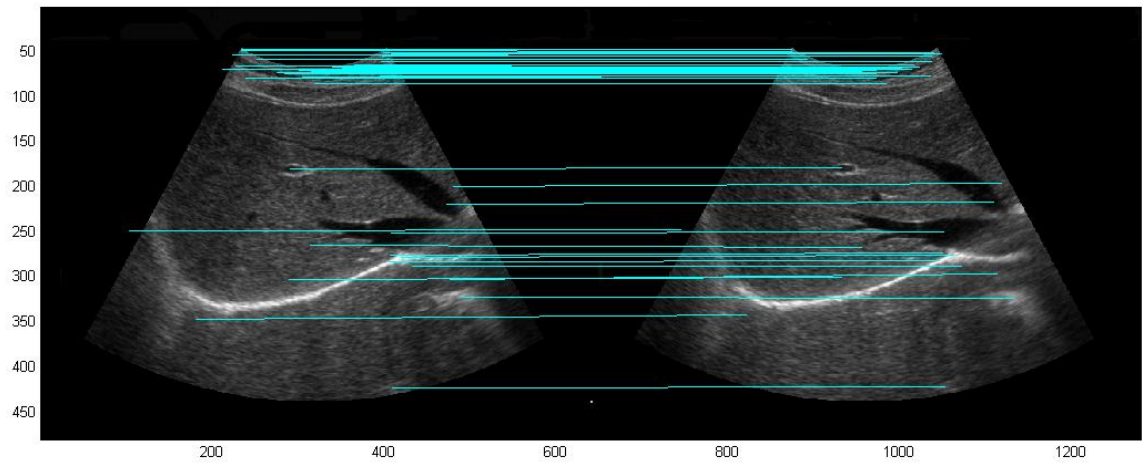
A. Invariant한 feature를 생성하기 위해 많은 과정을 거치는데(실시간에 적용하기에는 상당히 힘들것으로 판단됩니다.-640x480영상의 경우) 과정을 단순화 하면서 Invariant한 feature를 생성하게 할 수 있는 방법이 뭐가 있을까요? 꼭 영상에서 1000-2000개의 feature를 선택되게 해야 할까요? 1000개 미만이면 안될까요?

B. 제가 생각했을 때 SIFT가 상당히 강력한 Feature를 제공하고 있는 것 같기는 하지만 일부 응용의 경우(초음파 영상) 불필요한 과정이 있을 수도 있다는 생각이 듭니다. 이 부분을 개선해 볼 필요가 있을 것 같군요.

5. Questions and issues :

A. Match되는 Keypoint가 많을수록 좋은 것인가? 적은 Keypoint로 Match의 Accurate를 올릴 수 있는 방법은?

6. Experimental result :



- A. 왼쪽 영상의 Keypoint 906, 오른쪽 영상의 Keypoint 818, 정합된 Keypoint 55
- B. 초음파 영상에서처럼 노이즈가 심한 경우에는 매칭에 사용되는 Keypoint의 개수가 현저하게 적음을 알 수 있습니다. 이처럼 노이즈가 심한 경우 안정적으로 Keypoint를 찾을 수 있는 방법이 될까요? 여기서 안정적이란 노이즈에 강건한 Keypoint를 이야기 합니다. 실제 찾아진 Keypoint와 매치되는 Keypoint가 10%도 안된다는 것은 Keypoint를 선택할 때, 많은 개선이 있어야 할 것 같다는 생각입니다. 특히 노이즈가 심한 영상에서 말입니다.
- C. 같은 영상에서는 항상 같은 개수의 Keypoint가 찾아지는군요. 찾아지는 위치도 항상 일정한 오차 안에서 찾아 질까요? 매칭되는 Keypoint 역시 항상 같은 Keypoint가 매칭되는군요.