



고급 소프트웨어 실습(CSE4152)

SIFT 기반 이미지 매칭

목차

- 이미지 검색
- Scale Invariant Feature Transform (SIFT)
- 기타 응용 분야
- 실습: SIFT를 이용한 이미지 매칭 프로그램 구현
- 숙제: 특징점의 유사성 계산
- 참고문헌

이미지 검색

구글 이미지 검색 엔진
search by image
keyword

- 이미지 검색(image retrieval)은 주어진 검색어(keyword or metadata) 또는 이미지로부터 연관된 이미지를 찾는 것
- 기존에는 수동적으로 이미지에 할당하거나 웹페이지에서 얻어진 keywords를 통한 검색이 일반적으로 사용 되었지만 점차 이미지의 내용에 기반한 검색으로 발전 중
- Content Based Image Retrieval(CBIR)에서 주로 다루는 이미지 피처는 컬러, 텍스처(texture), 형태(shape), 객체(object) 등이 있음
- Scale Invariant Feature Transform(SIFT)은 CBIR 연구에서 오랫동안 사용된 방법으로 이미지 검색, 비디오 검색, 객체 추출 등 다양한 분야에서 이용됨



2000년경

Scale Invariant Feature Transformation (SIFT)

- SIFT 처리 기법은 이미지에서 회전, 크기 등의 변화에 영향을 받지 않는 특징점(keypoint)을 추출하고 그로부터 이미지의 변화에 강인한 디스크립터(descriptor)를 생성함
- 특징점과 디스크립터를 얻은 후에는 각 이미지 별로 유사한 특징점의 개수를 세어 이미지의 유사성을 판단

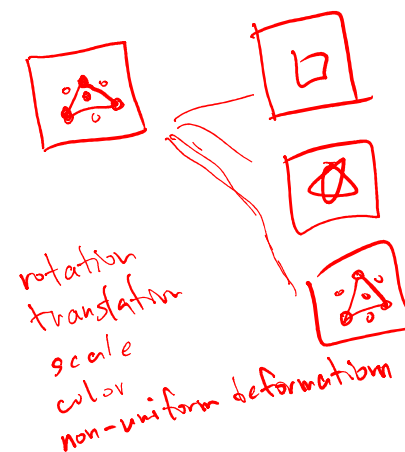
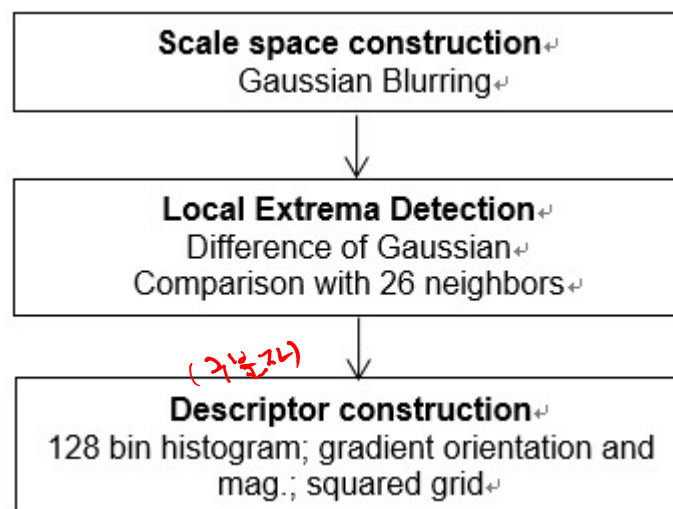
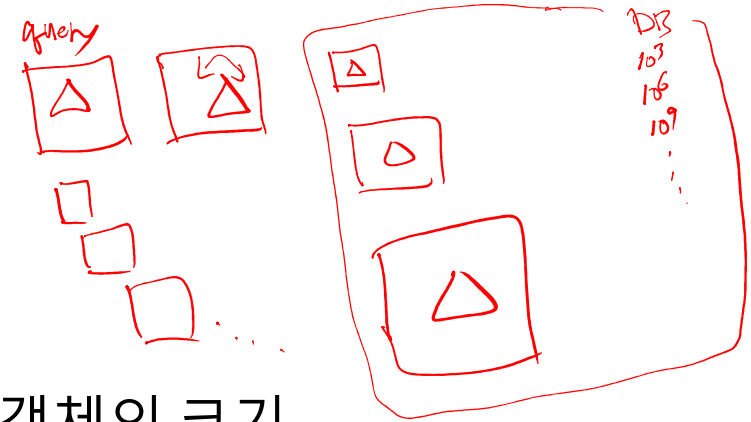
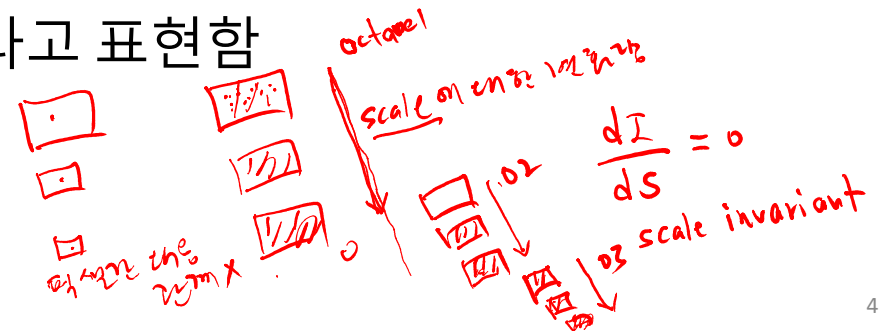
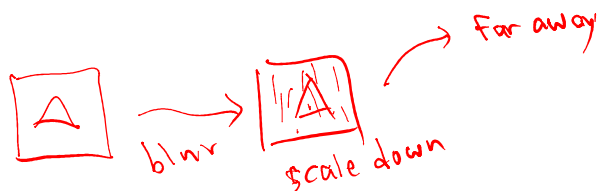


그림 1. SIFT 알고리즘의 도식적 표현

Scale space construction

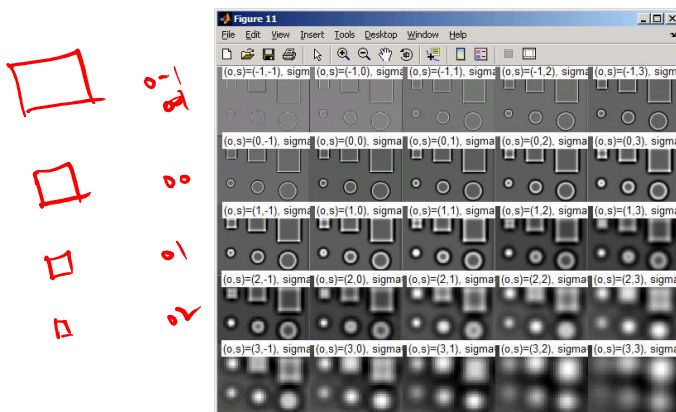
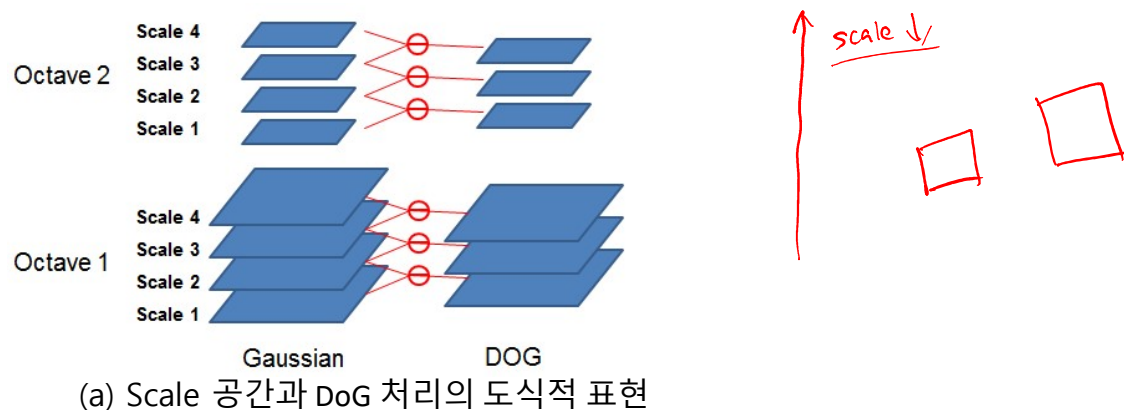


- scale 이란 이미지 또는 그 안에 나타나는 객체의 크기
- SIFT 알고리즘에서는 입력된 이미지를 인위적으로 여러 개의 scale 로 변환하고 그 안에서 관찰되는 특징점을 찾아냄으로써 크기 변화 문제 해결
- 입력된 이미지로부터 여러 개의 스케일로 표현되는 이미지를 생성하기 위하여 직접적으로 이미지의 사이즈를 줄이기 보다는 Gaussian blurring을 통해 scale 변화 효과를 달성
- scale 공간의 모사에는 한계가 있으므로 인위적인 이미지의 크기 변환도 병행하여 처리되며, 이는 Octave라고 표현함

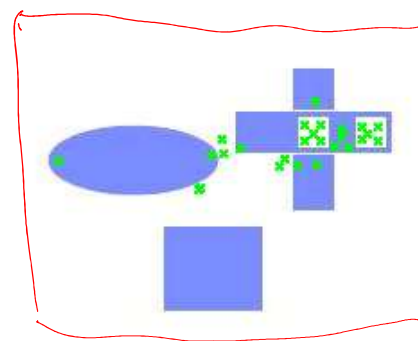


Scale space construction

- 각 이미지 당 여러 개의 Octave를 구성하고 각 Octave 안에 여러 개의 scale 공간을 형성



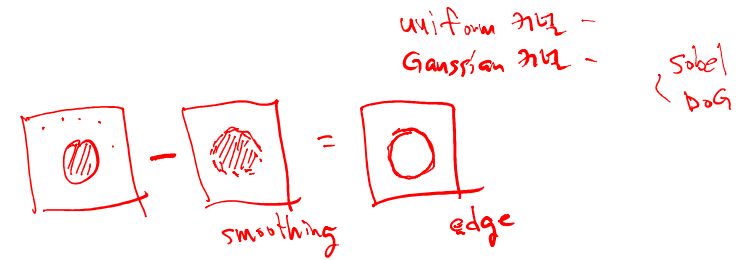
(b) 한 이미지를 Scale 공간에 표현한 예



(c) 간단한 이미지에서 얻어진 특징점

그림 2. Scale 공간의 도식적 표현과 SIFT 특징점의 예

Difference of Gaussian



- DoG는 보통 에지 검출용으로 쓰이던 방법이며, SIFT에서는 scale 공간에서 픽셀 값의 변화량을 찾기 위해 사용됨
- DoG에서 강한 반응을 보이는 픽셀은 Scale 공간에서 변화량이 큰 값

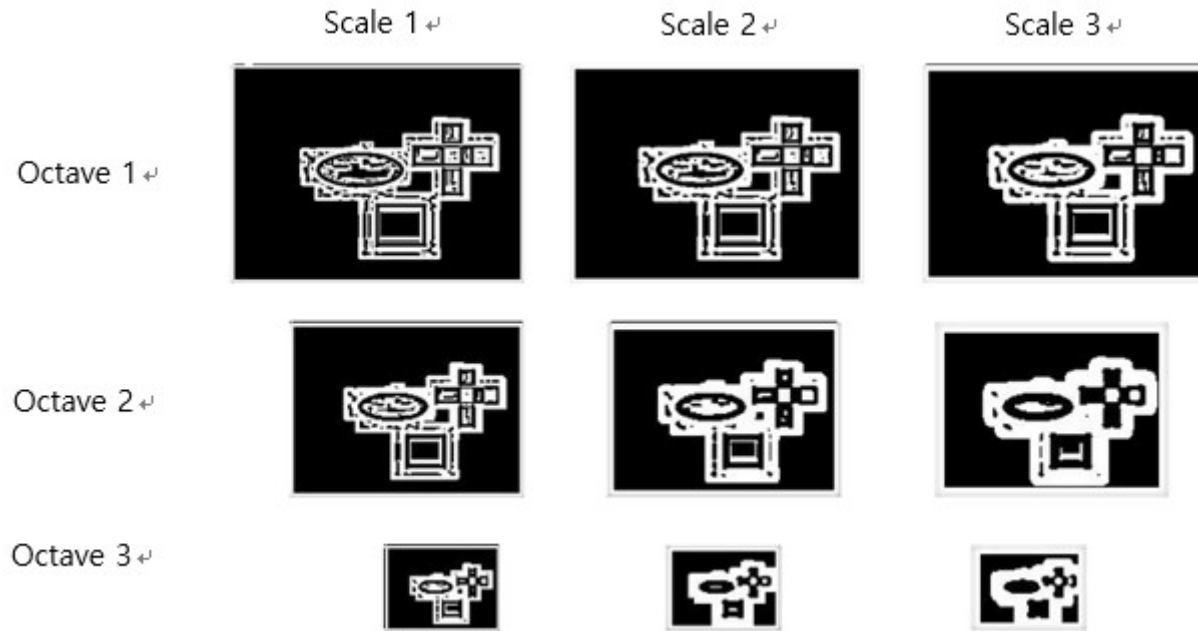


그림 3. 한 이미지를 Difference of Gaussian으로 표현한 예

Local Extrema Detection

- DoG로부터 특징점(Local Extrema, keypoint)을 찾는 방법은 SIFT에서 약간 독특하게 제안됨
- 스케일 공간에서 각 픽셀을 중심으로 주변 26개 픽셀과 비교하여 항상 크거나(최대) 항상 작으면(최소) 극값으로 선택
- 이렇게 얻어진 특징점들은 이후 몇가지 후처리 과정을 거치면서 이미지의 변화에 좀더 강인한 것들만을 추려냄

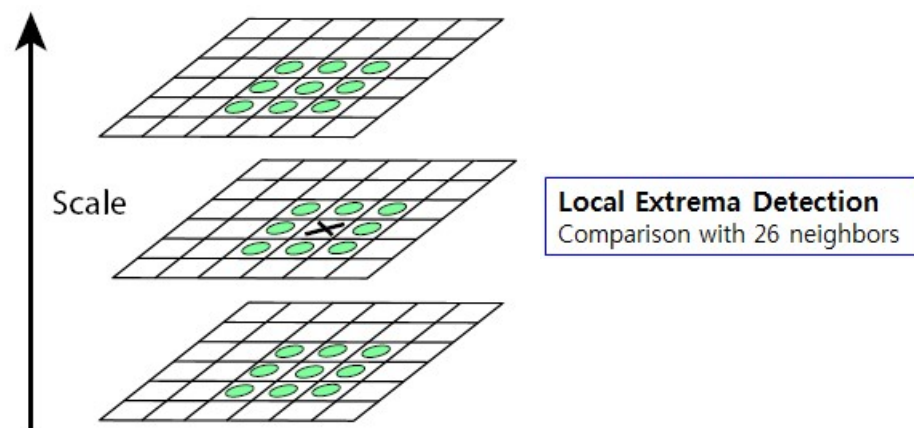
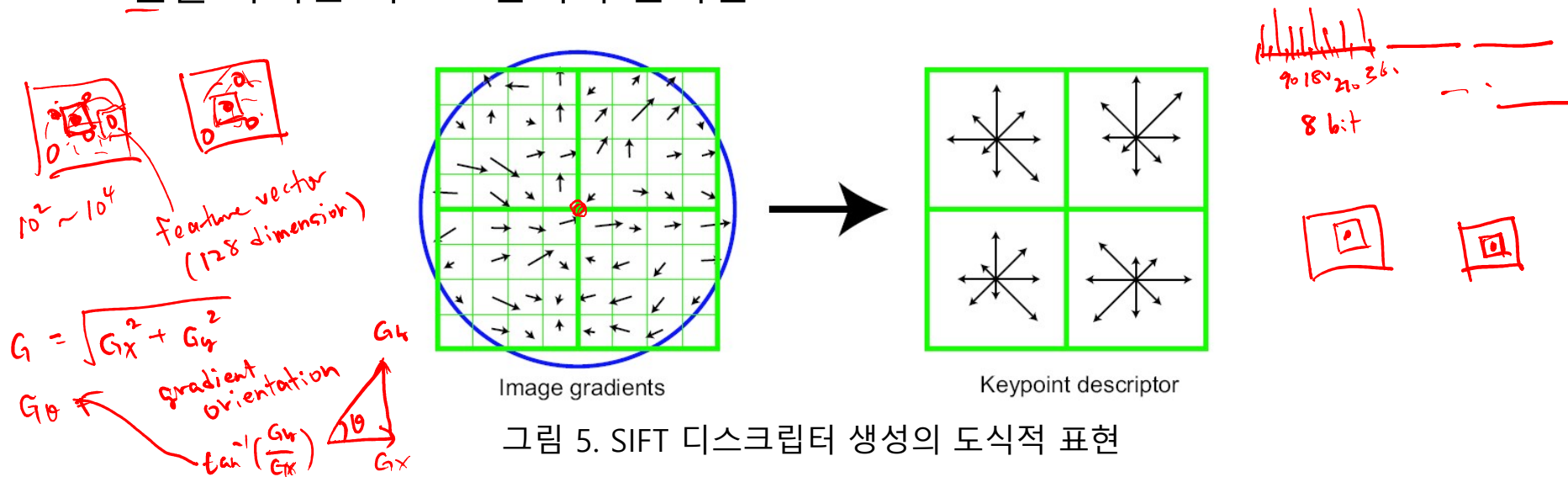


그림 4. 특징점 추출의 도식적 표현

Descriptor construction

- 특징점이 추출된 뒤에는 각 특징점 주변을 특정 길이의 디스크립터로 표현
- 특징점의 주변에 특정 크기의 윈도우를 정하고, 그 것을 4x4 영역으로 나누고 gradient 방향을 계산
- 단위 영역에서 얻어지는 Gradient(Week 2 실습 참고)의 방향을 8개의 정량화된 히스토그램으로 표현하고(45° 단위) 이것들을 이어 붙이면 8x4x4=128 차원을 가지는 디스크립터가 얻어짐

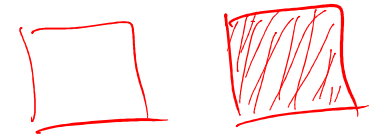


이미지 매칭

- 최종적으로 이미지의 유사성은 유사한 특징점의 숫자로 판단
- 특징점은 128 차원의 벡터로 표현되므로 L2 거리를 이용하여 공식 1과 같이 유사성을 측정

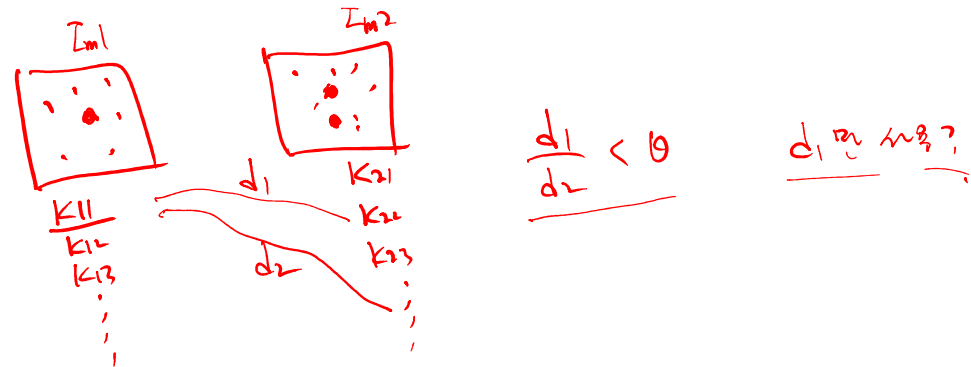
$$d = \sqrt{\sum_{i=1}^{128} (k_{1i} - k_{2i})^2}$$

식 1. L2-norm을 이용한 특징점 거리 측정

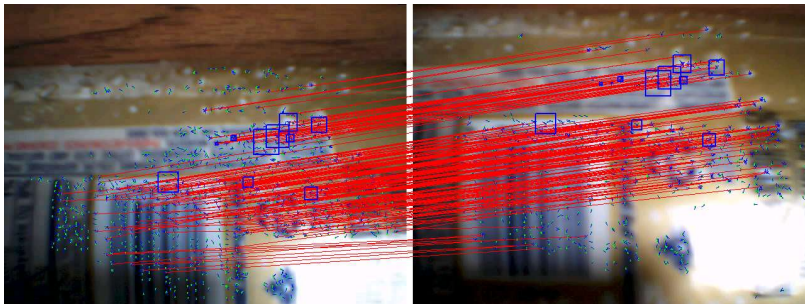


- 특징점들의 거리를 측정한 후 특정 threshold 값 이하를 매칭이라고 할 수도 있으나, 여기에는 이미지의 특성에 따라 다양하게 나타나는 특징점들의 거리를 일률적으로 처리하게 되는 문제점 존재

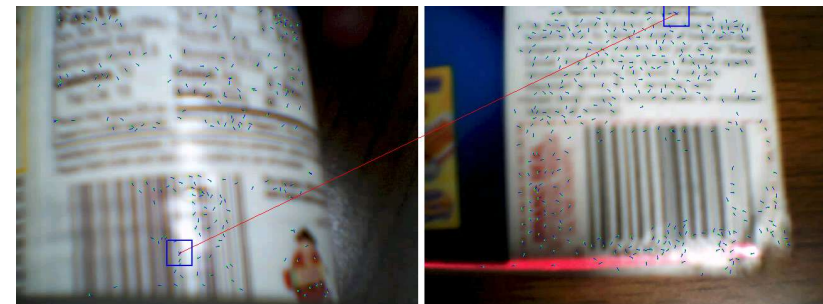
이미지 매칭



- 따라서 SIFT 알고리즘에서는 유사성이 아니라, 유일성(uniqueness)에 중점을 두고 특징점의 매칭 여부를 결정. 첫번째 이미지에서 얻어진 하나의 특징점에 대하여 두번째 이미지의 모든 특징점들과의 거리를 구하고 가장 가까운 거리 d_1 과 그 다음 가까운 거리 d_2 를 구하여 d_1/d_2 가 충분히 작다면 (일반적으로 0.49 사용) 매칭으로 간주



같은 물체의 경우 많은 특징점이 매칭됨

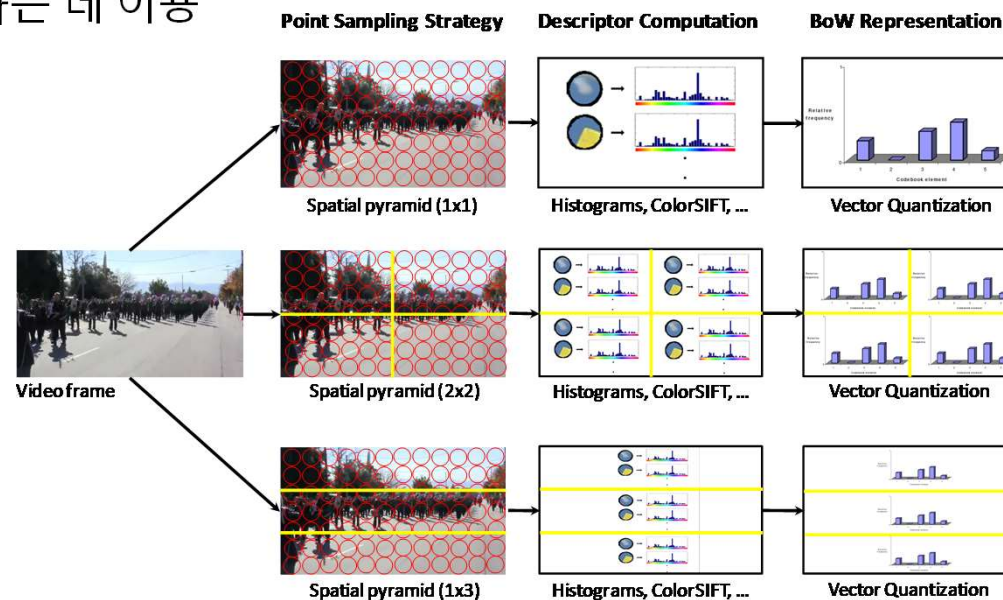


다른 물체의 경우 소수의 특징점이 매칭됨

그림 6. SIFT 매칭의 예

기타 응용 분야

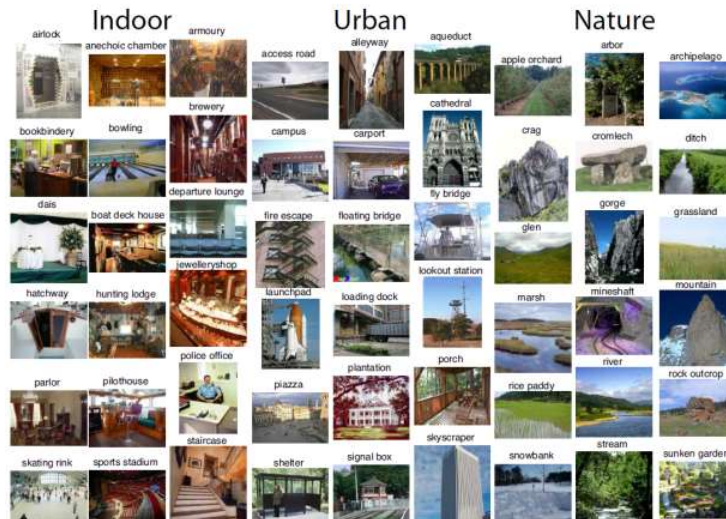
- SIFT는 특징점 추출 방식과 디스크립터로 특징점을 표현함에 있어서의 강인함으로 인하여 이미지 및 비디오의 처리에 광범위하게 이용되고 있으며 최근에는 생체 인식 분야의 얼굴 또는 지문 인식에도 이용
- 특징점 기반의 SIFT 매칭은 연산 속도가 느리므로 종종 Bag-of-words 방식과 결합되어 각각의 이미지 혹은 비디오를 고정된 길이의 피쳐 벡터로 나타내고, 이를 이용하여 이미지 및 비디오를 검색하는 데 이용



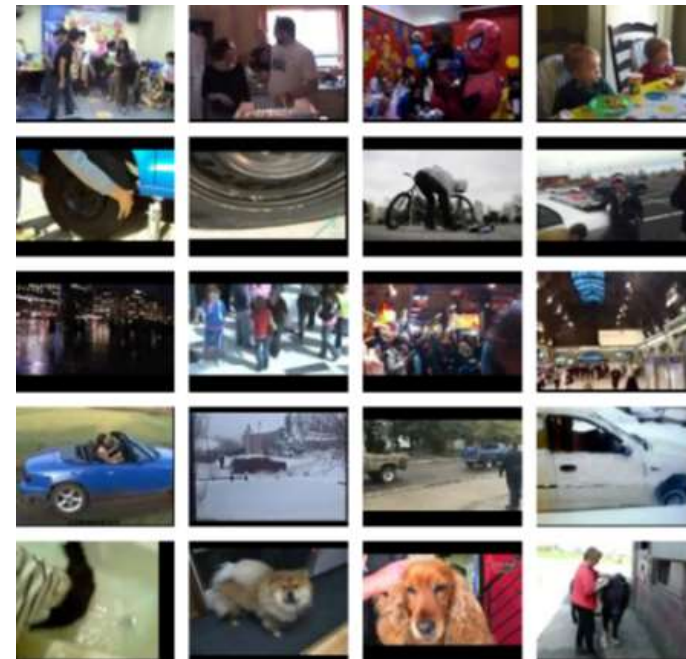
(a) Dense SIFT 특징점 추출 및 Bag-of-words 모델의 도식적 표현

기타 응용 분야

- Bag-of-words 모델은 방대한 이미지에서 추출된 SIFT 디스크립터들을 클러스터링하여, 자주 나타나는 대표적인 몇 개의 클러스터를 이용하여 각 이미지의 디스크립터들을 클러스터의 히스토그램으로 표현하는 방식



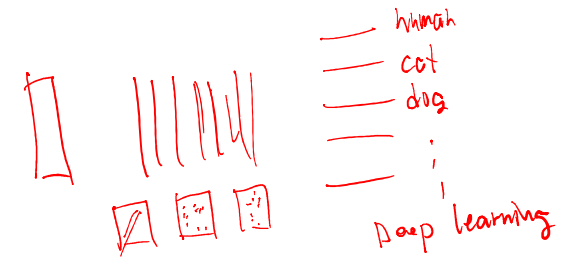
(b) 이미지 데이터 베이스



(c) YouTube 비디오 데이터 베이스

그림 7. Bag-of-words 모델의 도식과 그것이 주로 적용되는 이미지 및 비디오 데이터 베이스

기타 응용 분야



- Gaussian smoothing, local extrema 검출, descriptor 생성은 모두 사람에 의하여 고안된 알고리즘이 적용된 방식으로, 최근 영상 처리 연구의 주요 동향인 딥러닝을 적용하는 방법과는 차이가 있음
- 이미지 분류, 객체 검출, 얼굴 인식, 행동 인식 등 많은 이미지 분석 문제들에서 사용되던 기존 영상 처리 방법들은 최근 몇 년 간 모두 딥러닝 기법에 의하여 대체되었으나 특징점 검출 문제는 여전히 딥러닝 방법이 SIFT의 성능을 추월하지 못하고 있음

hand crafted feature
method



Classification



Detection



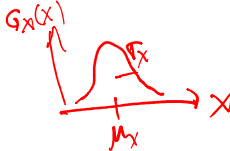
Segmentation

그림 8. 이미지 분류, 객체 검출, 객체 분할의 예

실습

1	4	7	4	1
4	16	26	16	4
7	26	41	26	7
4	16	26	16	4
1	4	7	4	1

$$\frac{1}{273} \times$$

$$h(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu_x)^2 + (y-\mu_y)^2}{2\sigma^2}}$$


X, Y 배열
($\sum h(x, y) = 1$) //

• Scale space 구성

- Scale space 구성을 위해 사용되는 cvSmooth 함수와 같은 기능을 수행하는 사용자 정의 함수를 구현, 위의 간략화된 2D 가우시안 공식을 사용하여 가우시안 커널을 구성한다
- Lab2에서 구현한 convolution 연산 함수를 최대한 재사용하여 Gaussian smoothing 함수 구성

• SIFT를 이용한 이미지 매칭

- 프로그램에 주어진 두 개의 함수는 각각 Image1-Image2, Image1-Image3 간에 SIFT를 이용한 특징점 매칭을 통해 이미지 간의 유사성을 측정한다
- 특징점 간의 유사성을 계산하고 이를 통해 이미지 간의 유사성 비교

숙제: 특징점의 유사성 계산

- 특징점 기반 이미지 매칭에서 설명한 $d1/d2$ 를 이용한 특징점 매칭의 문제점과, 이를 개선할 수 있는 방법에 관하여 설명한다.

참고 문헌

- M. Petrou and C. Petrou, Image Processing: The Fundamentals, 2nd ed., Wiley.
- R. Gonzalez, R. E. Woods and S. L. Eddins, Digital Image Processing Using MATLAB, 2nd ed., Gatesmark Publishing.
- David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004.