

CS231n Summary - Lecture 1

CV팀 임연우

Computer Vision이란?

Visual data에 대한 공부.

매우 방대한 양의 Visual data

Visual data 양이 터무니없을 정도로 매우 많음.

2015년도에 Internet traffic의 80%가 video라고함. 즉, 인터넷에서 전송되는 bit의 대다수가 visual data.

유튜브에 1초에 5시간 분량의 영상이 업로드됨.

그러므로 visual data를 활용하고 이해할 수 있는 알고리즘을 만드는 것이 중요함.

Visual data 활용에서의 문제점

하지만 "visual data는 이해하기가 힘들다"는 큰 문제점이 있음.

Visual data가 인터넷에 날라다니는 bits의 대부분을 차지하지만, visual data를 이루고 있는 것을 이해하는 알고리즘이 어려움.

물리학에서 우주에 있는 물질을 dark matter (암흑물질)이라고 하듯, 컴퓨터과학에서 visual data를 dark matter라고 부름. 우주의 많은 부분을 차지하고 존재하는 것을 알지만 관찰하기 어려운 암흑 물질처럼 visual data도 그러함.

Computer Vision과 연관된 다양한 분야들

Computer vision은 물리학, 생물학, 심리학, 컴퓨터과학, 공학, 수학 등등을 다 다룸.

History of Computer Vision

-Biological Vision

evolution's Big Bang : 540million years ago, 10million year이라는 짧은 기간 동안 동물 종이 폭발적으로 증가함. 후속연구에 의하면 처음으로 동물에게 "눈"이 생긴 시기.

즉 "vision이 시작된 것"이 종의 폭발적 증가를 초래함. (포식자는 피식자를 쫓고 피식자는 도망다니기 시작하면서 진화가 촉진됨)

Vision은 동물에게, 특히 인간에게 중요함.

인간은 피질에 있는 뉴런의 50%정도가 visual processing에 관여함. 가장 큰 감각 시스템임.

-Mechanical Vision, Camera

최초의 카메라는 1600년대 르네상스 시기의 obscura. Pinhole 카메라 이론에 기반함. 이는 초기의 동물 눈과 매우 유사함.

1950년대

Vision mechanism 연구 중 가장 영향력 있던 연구는, 1960년대 전기생리학 연구.

시각에 따른 고양이의 뇌의 전기신호 변화 관찰. "oriented edge에 반응하는 single cell이 있다는 것을 알아냄.

즉, visual processing이 시각세계에서의 simple structure에서 시작되고, complex visual world를 인식할 때까지, 뇌가 complexity를 build up해나간다는 것을 알아냄.

"시각 처리는 edges와 같은 단순한 구조로 시작되어 점점 복잡한 요소들을 처리하고, 궁극적으로 실제 시각적 input을 인지할 수 있게 된다."

1966년 MIT 여름 방학 연구

여름안에 대부분의 시각 체계를 구현하고자 했음. 이 작은 여름 프로젝트로 시작한 것이 아직까지도 현재의 computer vision분야 연구로 이어지는 중. 아직까지도 vision에 대해 다 알지 못하지만 빠른 속도로 연구 중.

1970년대

a. David Marr의 책 **VISION 저술. 컴퓨터 비전의 기본서**

컴퓨터가 사진을 보고 final holistic full 3d representation of the visual world에 도달하려면, 다음의 process를 거쳐야한다.

1) Primal sketch

edge, bar, end virtual line 등을 스케치

단순한 것으로부터 빌드업한다는 신경과학자들의 말과 유사함

2) 2와 1/2-D sketch

표면, 깊이, 층, 불연속성 등을 합쳐서 파악

3) 3-D sketch

이러한 vision에 대한 생각이 몇십년간 지배적이었음.

b) “how can we move beyond the simple block world and start recognizing or representing real world objects?”

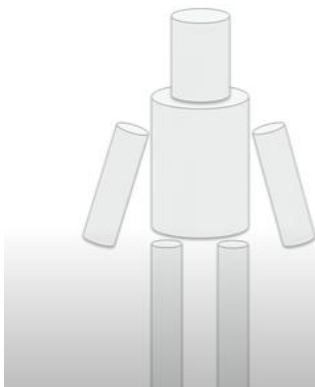
모든 object는 단순한 모양이나 기하학적 구성으로 이루어져 있다는 두 이론 등장.

1) *Generalized Cylinder* : 실린더 형태로 구성되어있다!

2) *Pictorial Structure* : pictorial 구조로 구성되어있다!

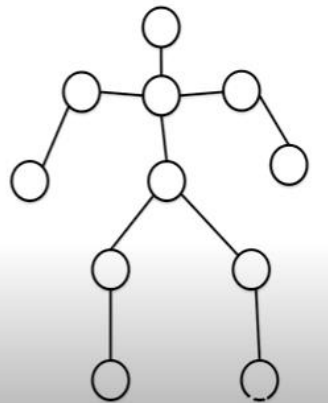
• Generalized Cylinder

Brooks & Binford, 1979



• Pictorial Structure

Fischler and Elschlager, 1973



객체를 단순한 기하학적 구성으로 이해해, 복잡한 객체를 단순화시키는 방법을 제안.

1980년대

David Lowe는 lines, edges, straight lines의 조합으로 단순화시키는 방법을 제안.



Object Segmentation의 시작

위의 많은 대담한 연구들 마저도 toy example에 불과. Real world에 적용하기에는 부족.

Object recognition이 어려우니까, Object segmentation에 집중하기 시작함.

(Object segmentation : 이미지의 각 픽셀을 의미있게 군집화하는 연구)

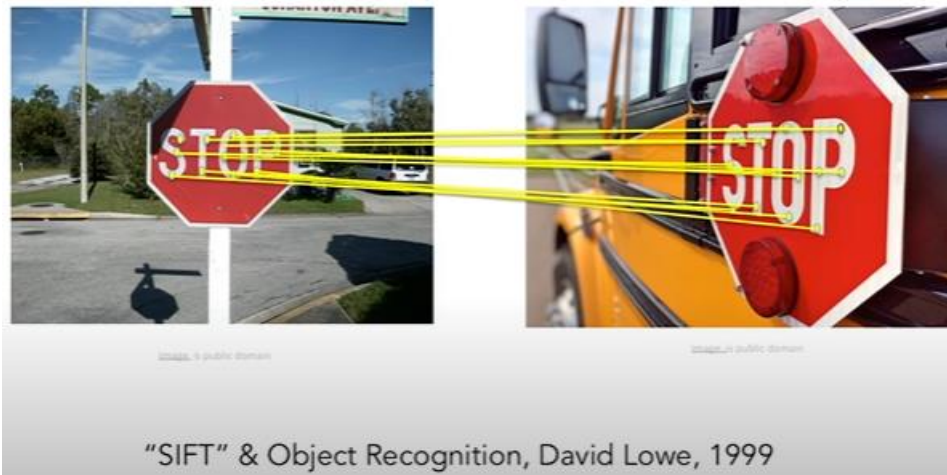
그룹지어진 것이 '사람'인지는 몰라도, 해당 픽셀들이 같은 의미를 가지는 것이라고 그룹화할 수는 있지 않을까? 라는 생각.

1990/2000년대

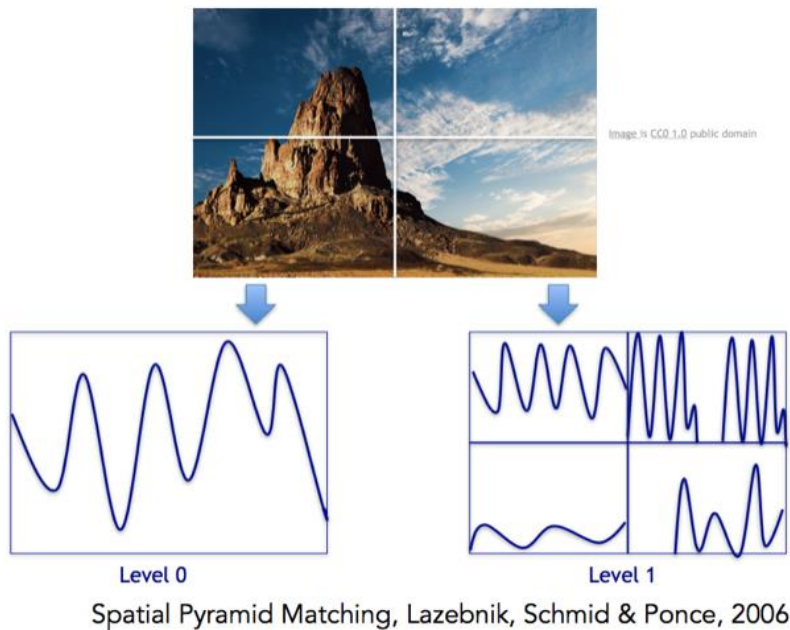
Machine learning이 탄력을 받으면서 Face detection(얼굴인식)이 가능해짐.

따라서 vector machines, boosting, graphical model, Neural Network의 첫 시작 등이 가능해짐.

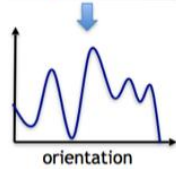
- **2001년** Paul Viola, Michael Jones: 실시간 얼굴인식 성공
- **2006년** Fuji film: 실시간 얼굴인식 가능한 디지털 카메라 판매 -> 이론이 실생활에 굉장히 빨리 적용됨
- **1990년~2000년** feature based object recognition(SIFT feature)



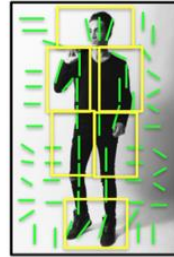
- 변화에 강하고 불편한 특징들을 발견해 매칭. 이미지 전체를 매칭하던 것에 비해, 중요한 특징들을 찾아 매칭 시키는 것이 더 쉬움.



이미지에서 여러 부분에서 피쳐를 뽑아내고 피쳐 디스크립터 (feature descriptor)에서 합침.
서포트 벡터 머신 (SVM)알고리즘을 그 위에서 수행.



Histogram of Gradients (HoG)
Dalal & Triggs, 2005

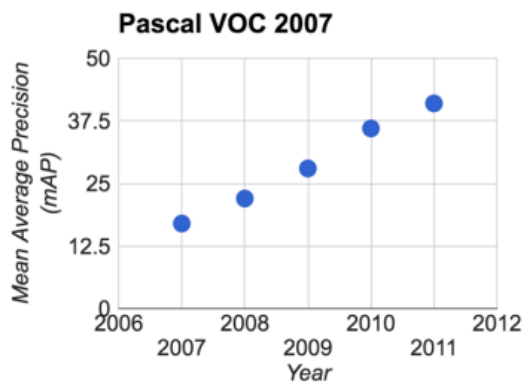


Deformable Part Model
Felzenszwalb, McAllester, Ramanan, 2009

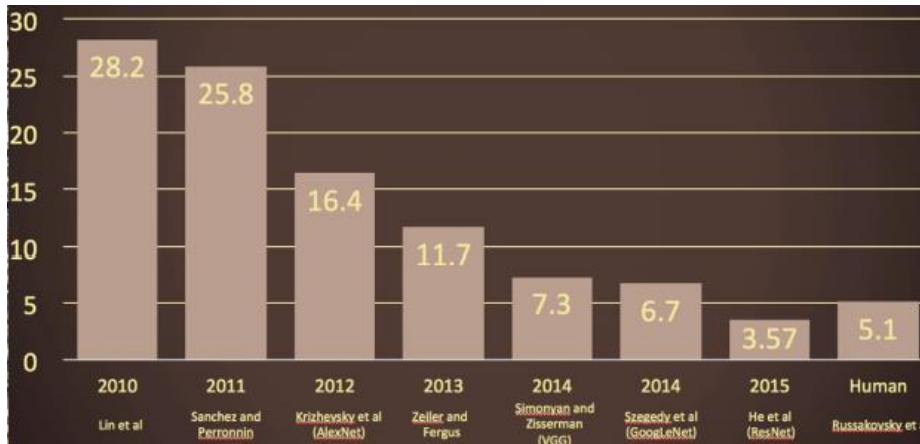
21세기

인터넷과 카메라의 발전으로 실험데이터들의 질이 급격히 상승. 이때부터 양질의 데이터셋을 모으기 위한 움직임이 이어짐.

PASCAL Visual Object Challenge(VOC): 알고리즘 테스트에 사용되었고, 밑의 그래프는 2007년부터 2012까지의 객체인식 성능을 보여줌.



ImageNet: 가장 큰 데이터셋을 만들어 Overfitting 을 방지하고 일반화 능력을 키워 이 세상 모든 객체들을 인식할 준비. ILSVRC 대회를 열어 해당 데이터셋으로 지속적인 알고리즘 테스트를 진행함.



2012 년도의 CNN. 오차를 급격히 감소시킴. 이 수업에서는 CNN 을 중점적으로 다룸.

CNN 이 2012 년도에 갑자기 나타난 것은 아님. 이미 90 년대에 이미 LeNet 아키텍처가 존재. 하지만 90 년대에 비해 비교할 수 없을만큼의 데이터셋과 연산량의 증가가 오류율 감소에 큰 몫.

CS231n 강좌 overview

해당 수업에서 다룰 주제

- image classification : 여기에 초점 맞춤.
- object detection
- image captioning

CNN 으로 물체 인식을 잘 할 수 있게 됨. 합성곱 신경망인 Conv net 을 도입한 이후부터 크게 진전함.

컴퓨터 비전의 최종목표

사람처럼 볼 수 있는 기계를 만들자

더 연구해야 할 연구주제

Semantic Segmentation, Perceptual Grouping, 3D understanding, 행동인식