Intro to CV

Autopilot Full Self-Driving Hardware

https://vimeo.com/192179726

COCO Explorer: Show Captions

http://cocodataset.org/#explore

Google Vision API

https://cloud.google.com/vision

Computer Vision?

- **OCR (Optical Character Recognition):** 문서의 텍스트를 인식하고 식별.
- **Vision Biometrics :** 홍채 패턴 인식을 통해 사람들을 구분합니다.
- **Object Recognition :** 실시간의 이미지나 스캔 입력을 가지고 제품을 분류합니다.
- **Special Effects :** 모션 캡처 및 모양 캡처, 영화에서의 CGI.
- **3-D Printing and Image Capture :** 영화, 건축 구조 등에 사용됩니다.
- **Sports :** 경기에서 필드에 추가 라인을 그리거나 이를 기반으로 영상 판독을 합니다.
- **Social Media :** 얼굴 모양에 맞추어 착용되는 이미지.
- **Smart Cars :** 사물과 사람을 인지할 수 있습니다.
- **Medical Imaging :** 3D 이미징 및 이미지 유도 수술.

Marvin Minsky, MIT
Turing award,1969

방송용 카메라를 컴퓨터에 연결하고 보이는 것을 설명하도록 하는 프로젝트

언제: 1966년 여름 동안
누가: Minsky 교수 + Gerald Sussman + 10 여명의 학생그룹

실제 연구 내용에 대한 링크
https://dspace.mit.edu/handle/1721.1/6125
ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-100.pdf

참고: 컴퓨터 비전의 도시 전설:
http://www.lyndonhill.com/opinion-cvlegends.html

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Mission
“컴퓨터가 이미지를 이해하게 하는 것”

Visual Processing Mechanism

프로그램을 어떻게 만들어야 하나?

Rule Based Programming

vs

Machine Learning

사진에 무엇이 있는가?

무엇을 하고 있는 사진인가?

사진을 묘사할 수 있는가?

visual recognition problems

- object detection
- semantic segmentation
- image classification
- action classification
- captioning

Why is this hard?

243	239	240	225	205	185	188	218	211	206	216	225
242	239	218	110	67	31	34	152	213	206	208	221
243	242	123	58	94	82	132	77	103	208	208	215
235	217	115	212	243	236	247	130	51	209	208	211
233	208	131	222	219	226	156	114	74	208	213	214
232	217	131	116	77	150	69	56	52	201	228	223
237	237	182	186	184	179	159	129	69	237	235	235
232	236	201	154	215	133	129	81	175	252	241	240
235	238	230	128	172	138	65	63	234	249	241	245
237	236	247	143	56	78	10	94	255	248	247	251
234	237	245	193	55	33	115	144	213	255	253	251
248	245	161	128	143	109	138	65	47	156	239	255
100	107	39	102	94	73	114	58	17	7	51	137
23	32	33	148	168	203	179	43	27	17	12	8
17	26	12	160	255	255	109	22	26	19	35	24

Pixel (픽셀)

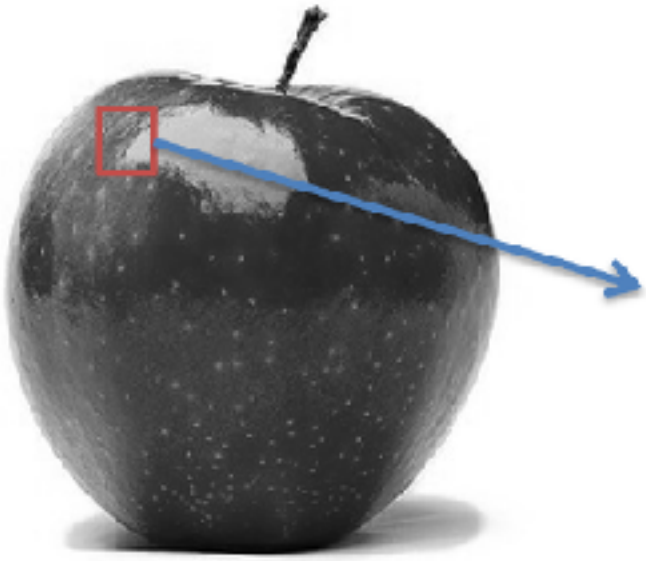
picture element - 화면을 구성하는 가장 기본이 되는 단위

표준 디스플레이 해상도는 크게 다음과 같다.

이름	해상도 (메가픽셀)	가로 x 세로
CGA	0.064	320×200
EGA	0.224	640×350
VGA	0.3	640×480
SVGA	0.5	800×600
XGA	0.8	1024×768
SXGA	1.3	1280×1024
UXGA	1.9	1600×1200

위키피디아 링크:https://ko.wikipedia.org/wiki/%ED%99%94%EC%86%8C

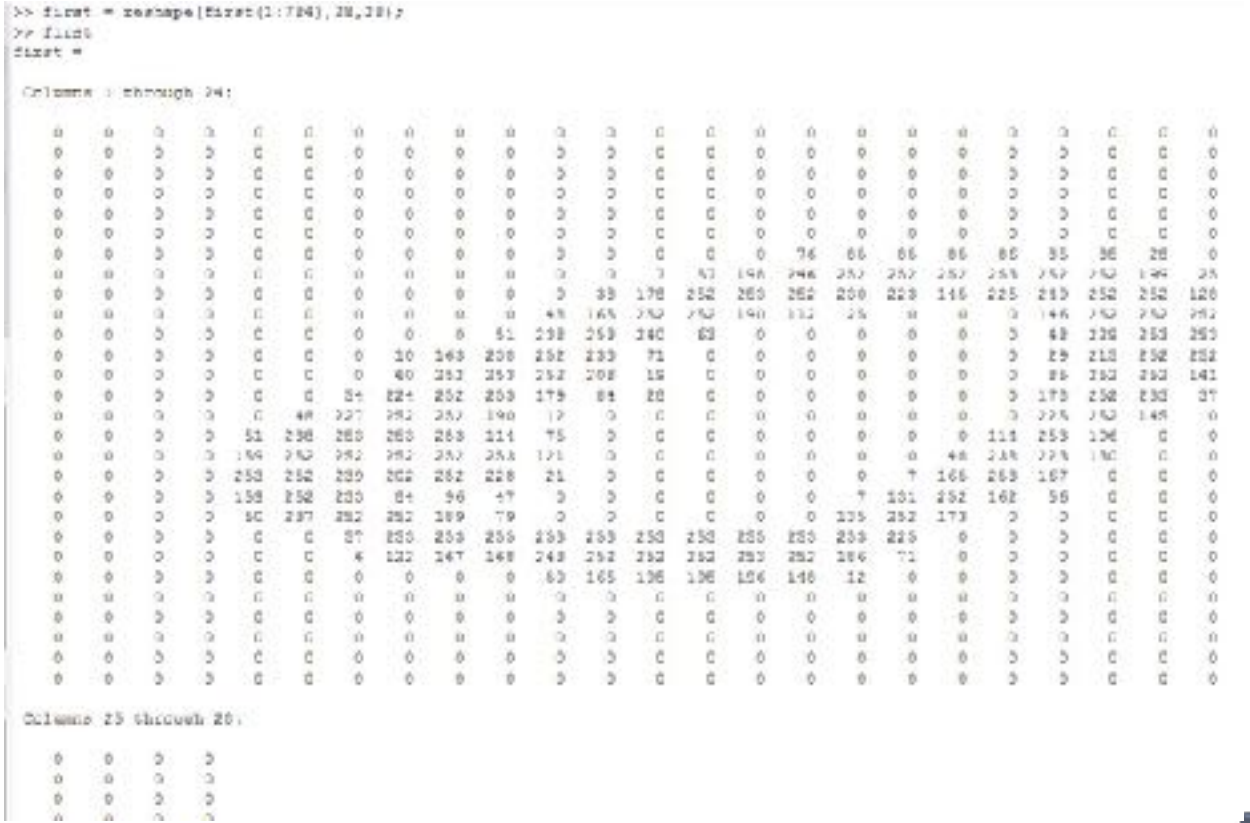
$$
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .6 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .5 & 1 & .4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .7 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .9 & 1 & .1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .3 & 1 & .1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$



Gray-scale Image

206	225	198	185	182	174	151	137	132	151	140	119	132	120	141	121	125
230	194	191	197	189	160	164	143	156	157	99	160	127	113	121	122	107
195	188	213	185	148	173	153	160	146	116	123	123	155	142	107	151	117
183	191	190	170	177	167	148	164	145	134	127	158	140	112	128	97	146
194	177	189	184	171	138	179	145	100	142	146	140	122	114	115	109	164
177	171	170	175	161	165	172	121	149	153	127	116	131	148	133	133	140
180	177	157	156	160	171	137	145	130	134	130	149	130	128	137	123	119
185	160	171	149	157	142	132	147	124	129	118	138	132	118	165	138	104
165	159	145	176	153	125	153	137	131	142	152	152	116	135	147	106	122
153	180	186	168	139	160	151	158	114	155	172	83	125	154	107	124	152
176	191	153	127	165	140	144	145	164	138	71	164	166	81	147	150	132
177	145	124	151	152	154	140	179	156	92	161	201	108	101	165	128	131
139	131	152	146	140	158	173	155	92	170	171	89	123	161	124	136	99
145	136	169	150	141	134	175	106	156	155	142	121	144	137	102	112	107
141	157	158	121	139	169	137	136	156	124	145	129	105	104	118	112	118
158	149	122	135	153	147	107	156	121	152	156	118	124	129	118	104	94
165	142	145	132	165	117	135	146	127	138	107	95	116	120	102	94	93
130	168	151	132	132	134	125	136	116	132	126	111	129	106	99	102	123
171	173	149	136	133	111	130	121	120	132	104	127	120	111	106	102	118
165	171	150	109	130	125	120	114	105	121	109	111	111	103	115	100	96
181	138	124	129	102	123	107	136	119	101	108	109	114	95	102	109	125
155	137	131	109	114	105	128	119	104	132	103	121	104	129	103	124	110
140	120	139	128	103	116	110	122	110	136	103	112	110	108	124	120	104
119	111	136	112	125	125	122	115	90	119	105	98	132	101	126	91	122
125	127	132	91	134	121	82	117	109	96	97	112	130	109	113	126	129

컴퓨터에게
경계(외곽선)를
가르쳐 봅시다.

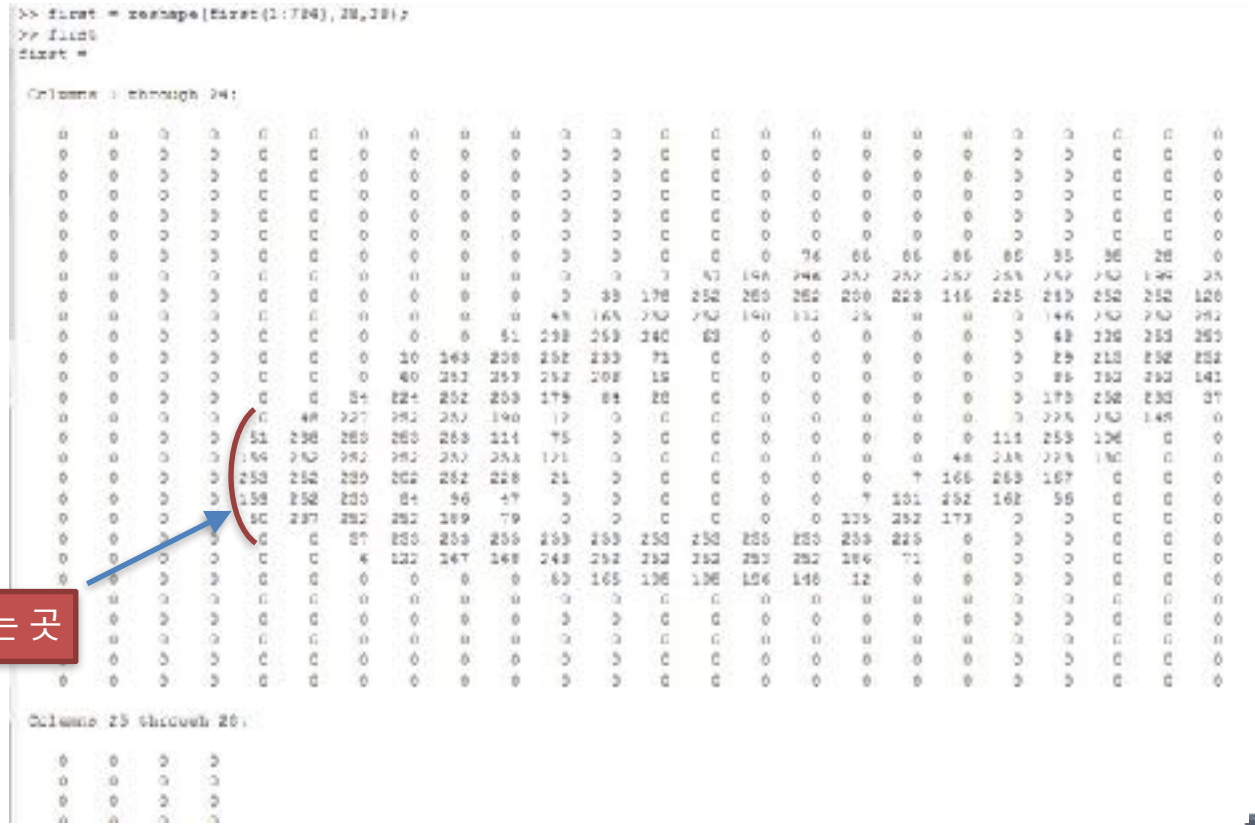


컴퓨터에게 경계(외곽선)를 가르쳐 봅시다.

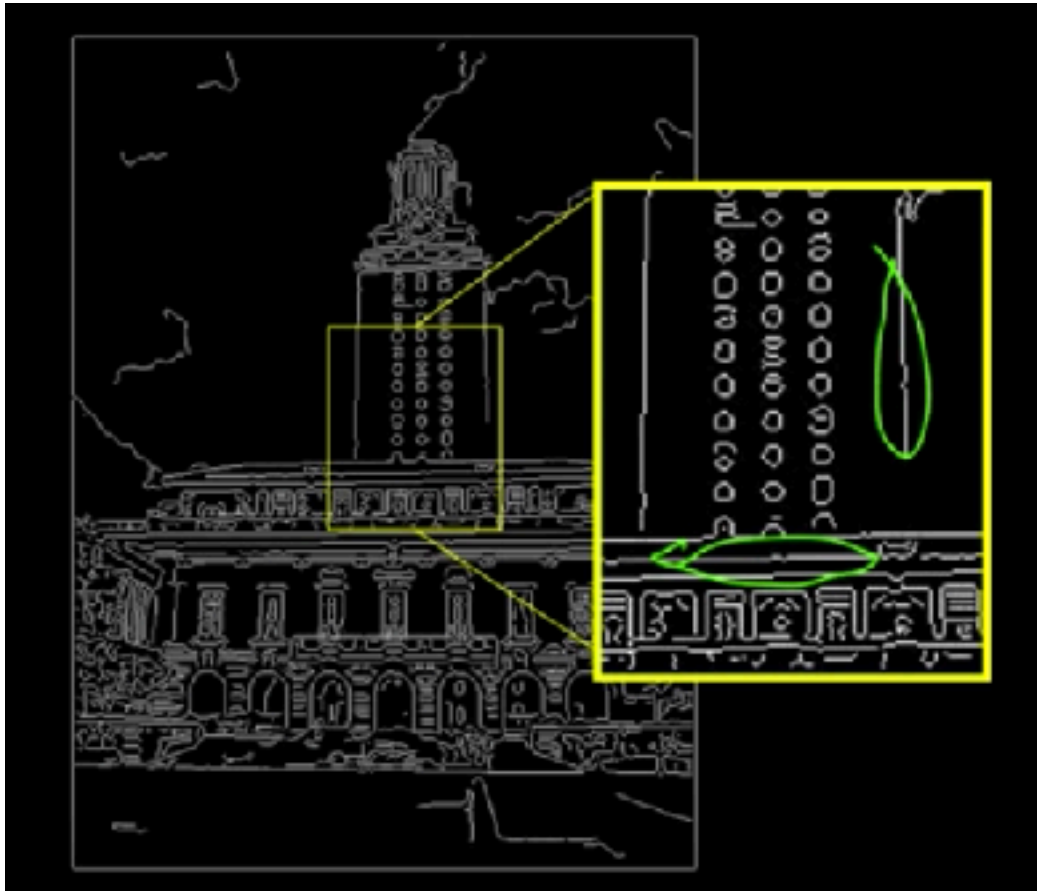
값의 급격히 변화하는 곳이 경계

미분을 이용하여
값이 급격히 변화하는 곳을
모두 찾아서 연결하면
경계를 찾을 수 있겠다!

x 방향으로 값이 변화하는 곳



이제 실제 이미지에서 경계를 찾자.

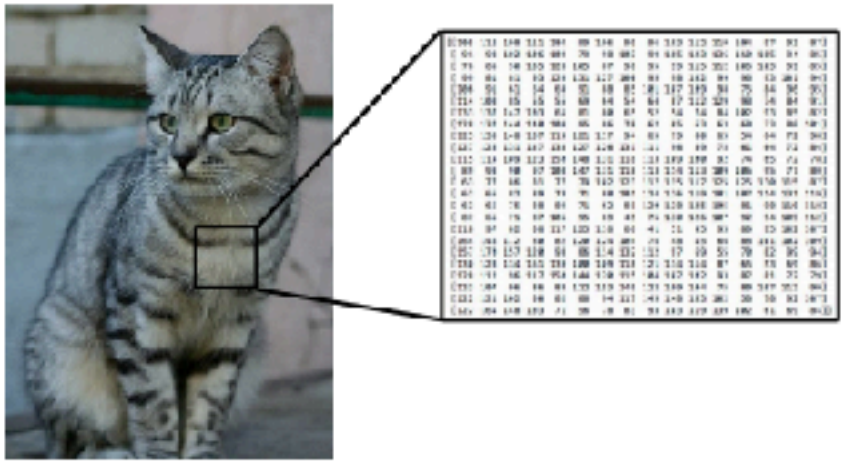


Mission Failed!
경계를 찾는 것에 실패하였습니다.

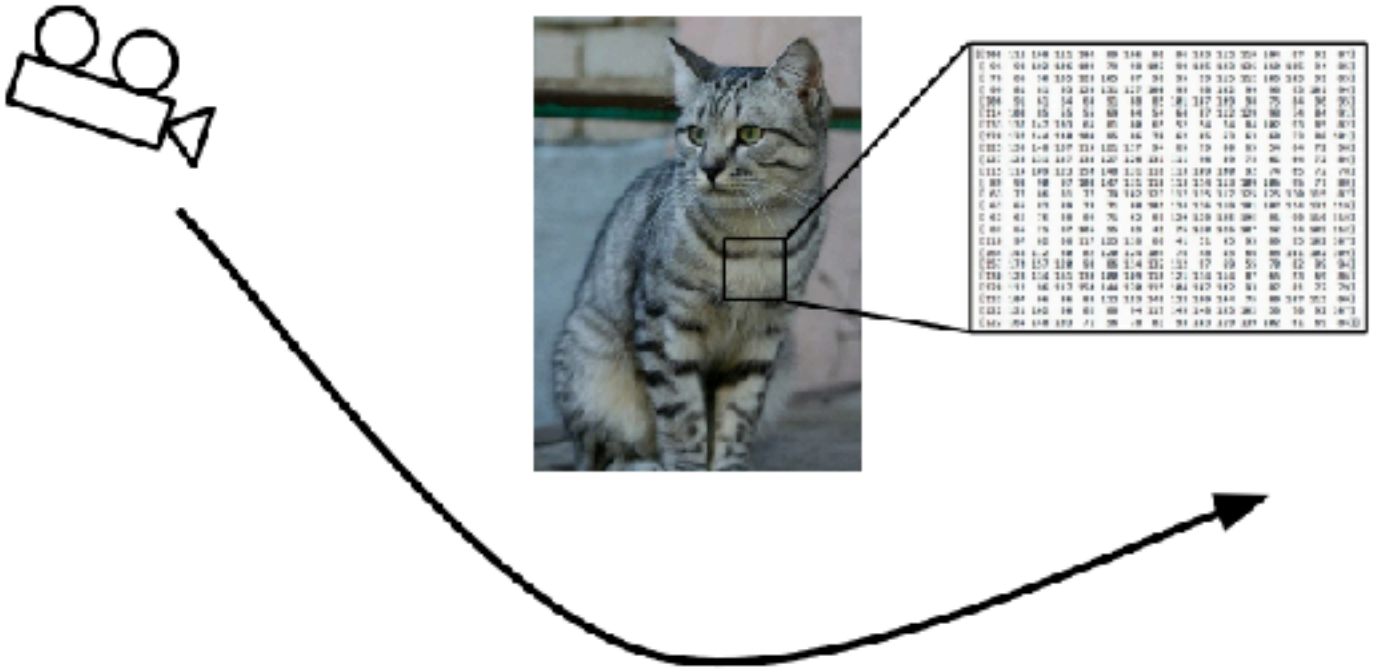
- 경계의 일부분이 제대로 찾아지지 않아요.
  (예외 상황이 천차만별)
- 찾은 경계에는 종종 노이즈가 존재해요.

Object Detection은 먼나라 이야기
Edge Detection 조차 쉽지 않아요!

어쨌든 컴퓨터에게 아래와 같은 모양을 찾아서 ‘고양이’라고 알려주었습니다.



같은 고양이라도 카메라 각도에 따라 이미지(데이터)가 달라지네

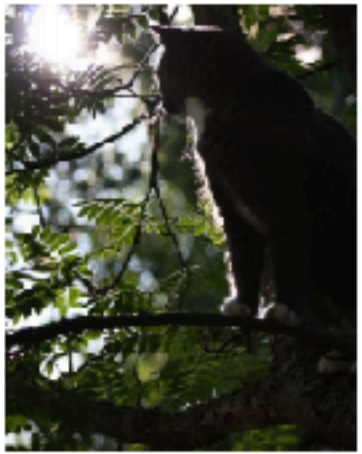


어라 빛에 따라서 다시 이미지(데이터)가 바뀌네

This image is CC0 1.0 public domain

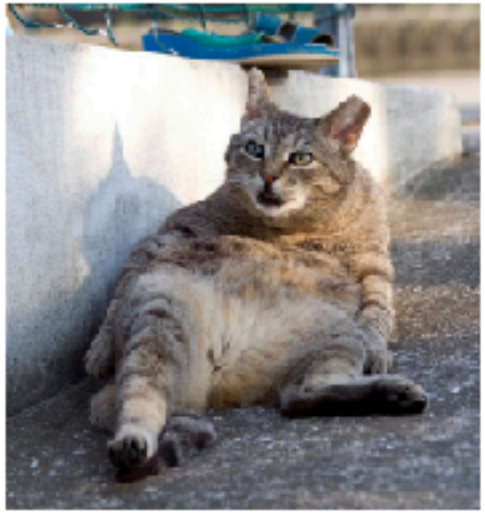
This image is CC0 1.0 public domain

This image is CC0 1.0 public domain

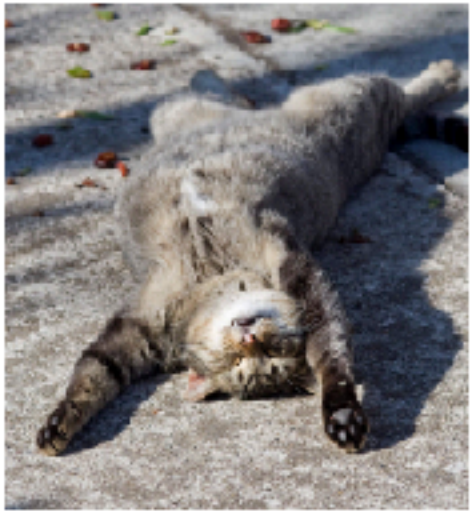


This image is CC0 1.0 public domain

고양이는 가만히 있지를 않구나...

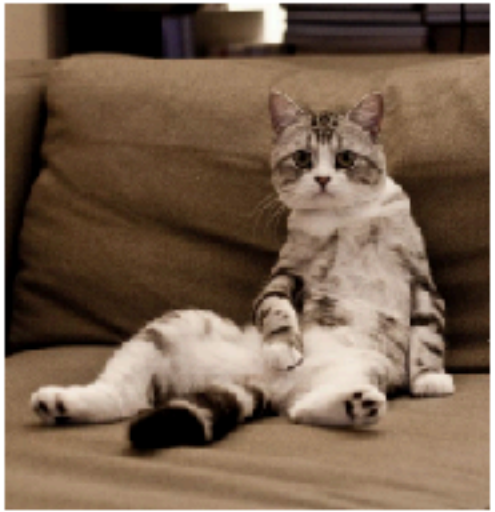


This image by Umberto Salvagnin
is licensed under CC-BY 2.0



This image by Umberto Salvagnin
is licensed under CC-BY 2.0

This image by sara.bear is
licensed under CC-BY 2.0



This image by Tom Thai is
licensed under CC-BY 2.0

어떤 물건에 가려질 수도 있고,

This image is CC0 1.0 public domain

This image is CC0 1.0 public domain

This image by janeb13 is licensed under CC-BY 2.0

배경의 색과 비슷한 경우도 있구나

This image is CC0 1.0 public domain

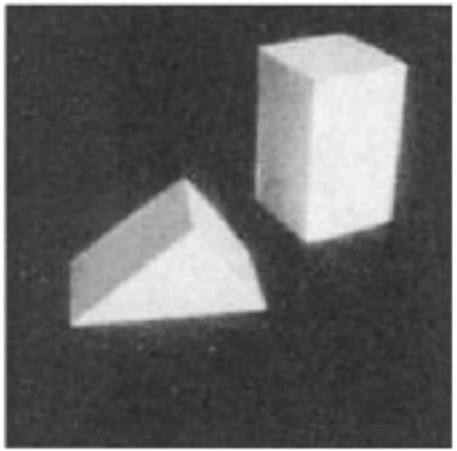
This image is CC0 1.0 public domain

다섯 마리라는 걸 어떻게 알려주지?

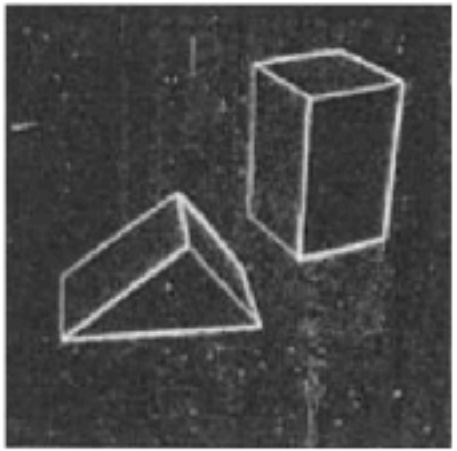

history
first
way
good
culture
calendar
historians
ancient
Egypt
civilization
great
fact
even
like
America
AD
Christianity
now
important
man
need
Rome
BC
one
modern
years
world
History
historical
new
human
year
make
Jesus
Western
event
anchor
Dark
Columbus
facts
Roman
life
know
chapter
Ages
story
just
religious
time
born

1960's - 간단한 Block으로 구성된 이미지를 해석함.

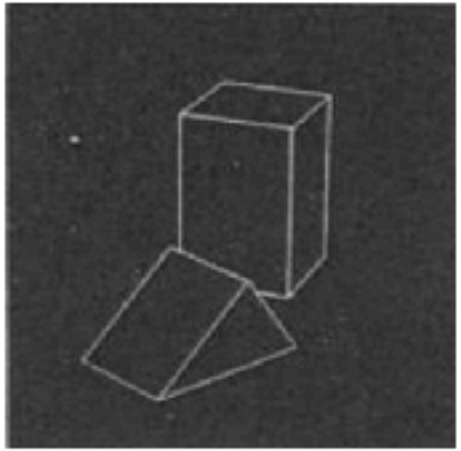
Larry Roberts
Father of Computer Vision"



Input image



2x2 gradient operator



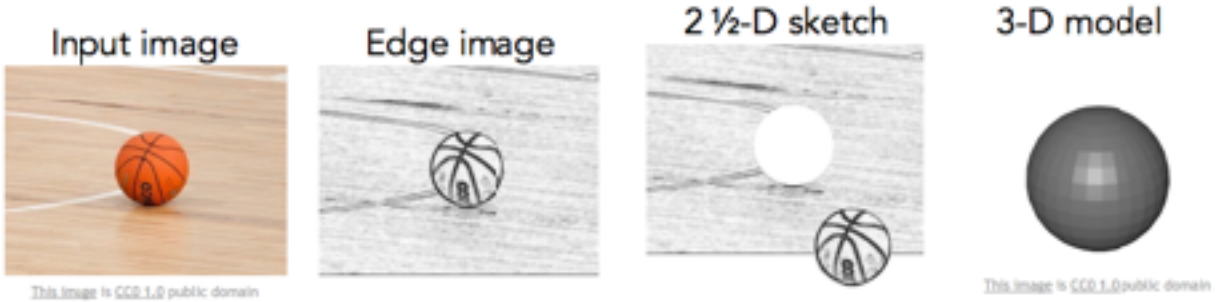
computed 3D model
rendered from new viewpoint

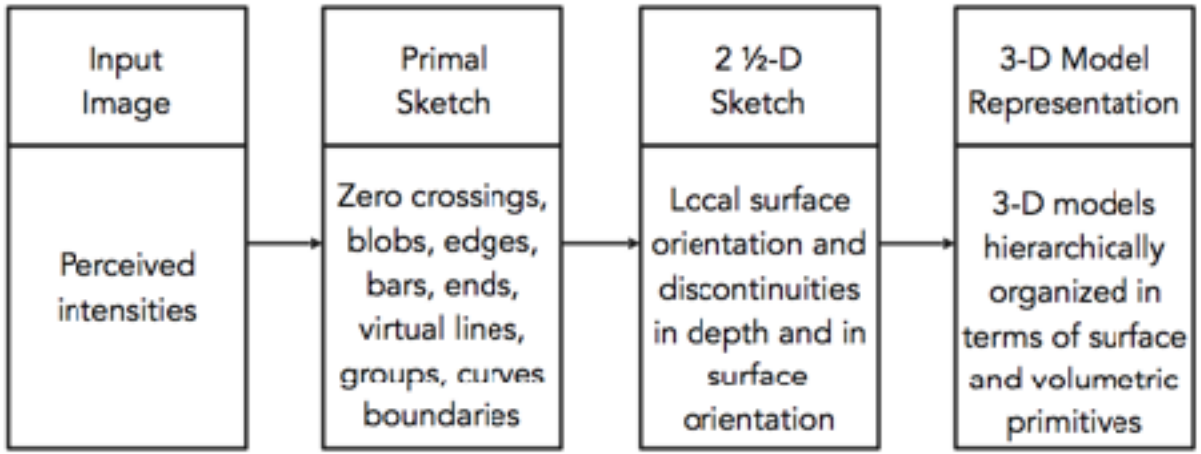
컴퓨터에게 “에지”를 이해시켰어요.
선과 면을 이해하게 되었지요.

하지만, 사물을 의미있는 수준으로
구분하는 것은 실패했어요.

ㅠㅠ

1970's - 이미지 선택 기술과 선택된 이미지의 해석 기술 진전.





Stages of Visual Representation, David Marr, 1970s

컴퓨터에게 사물간 “계층(Layer)”을 이해시켰어요.

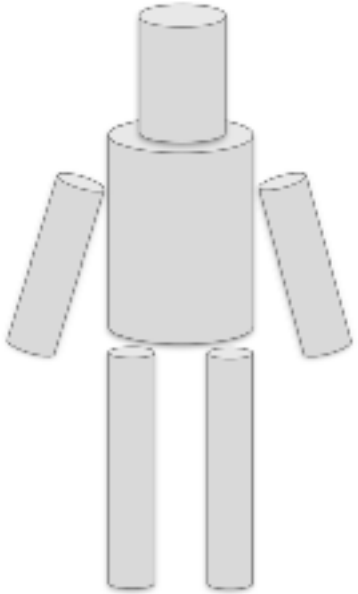
하지만, 계층을 이해할 수 없는
케이스가 너무 많았어요.

ㅠㅠ

1970's - 이미지 선택 기술과 선택된 이미지의 해석 기술 진전.

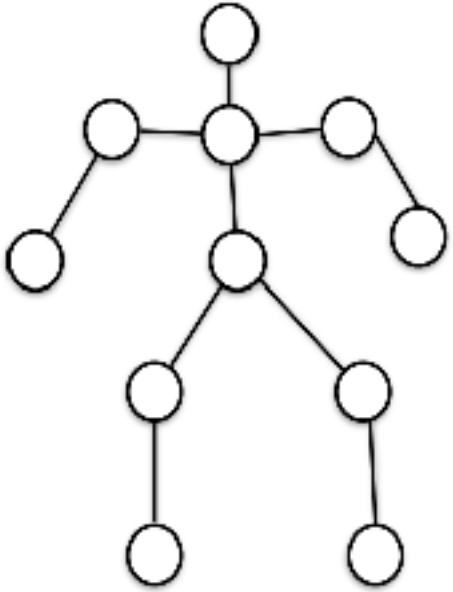
- Generalized Cylinder

Brooks & Binford, 1979



- Pictorial Structure

Fischler and Elschlager, 1973



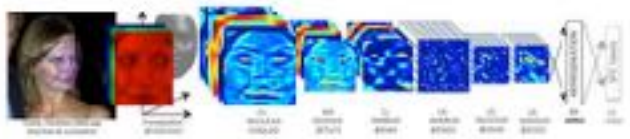
컴퓨터에게
사물의 움직임을 이해시켰어요.

여전히 사물을 의미있는 수준으로
구분하는 것은 못하고 있어요.

ㅠㅠ

1990's - 안면 인식 및 인식 기술 확장

Deep Face (Facebook)



Model	# of parameters	Accuracy, %
Deep Face Net	128M	97.35
Human level	N/A	97.5

Training data: 4M facial images

Y. Taigman, M. Yang, M.A. Ranzato, L. Wolf. DeepFace: Closing the Gap to Human-Level Performance in Face Verification // CVPR 2014.

"SIFT" & Object Recognition, David Lowe, 1999

컴퓨터가 사물의 스케일을 이해시켰어요.

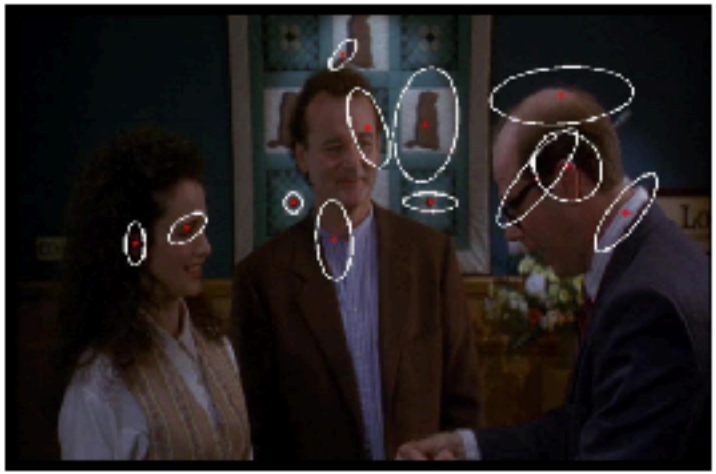
여전히 사물을 의미있는 수준으로
구분하는 것은 못하고 있어요.

ㅠㅠ

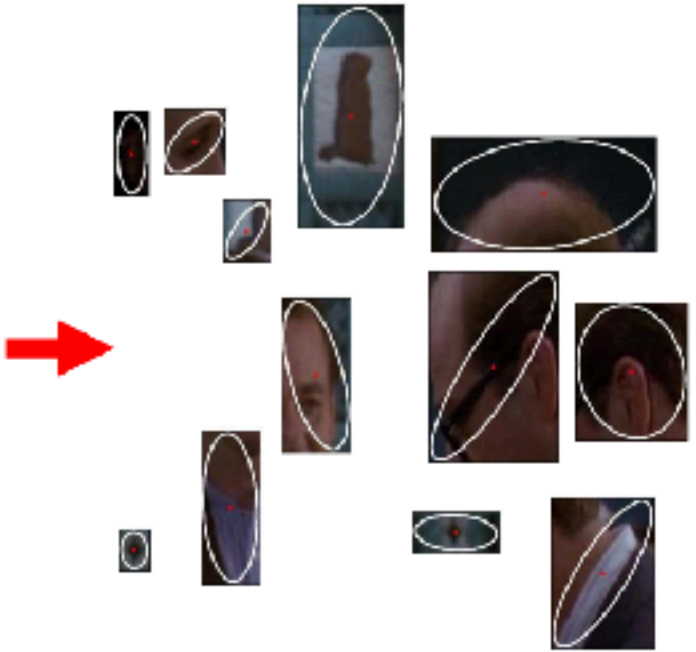
Representation: Bag of visual words

Visual words are 'iconic' image patches or fragments

- represent the frequency of word occurrence
- but not their position



Image



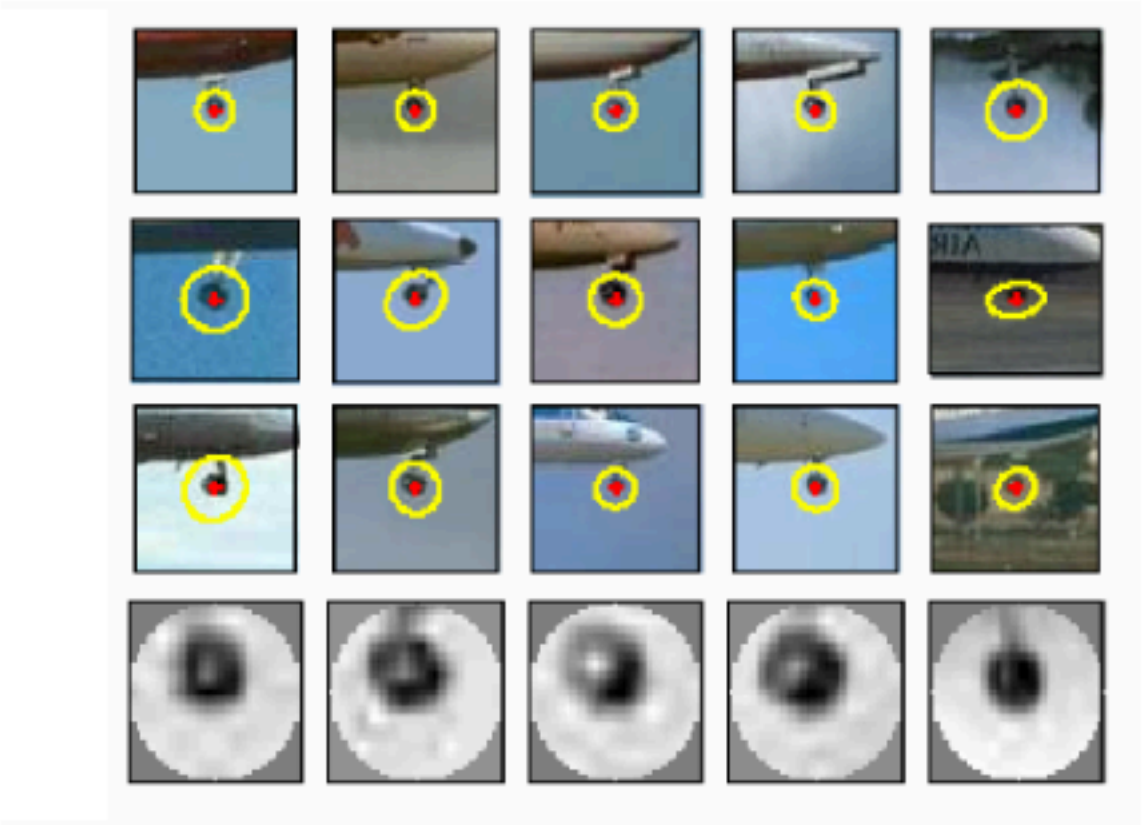
Collection of visual words

컴퓨터가 “특정 이미지 조각”을 가지고
분류할 수 있게 만들었어요.

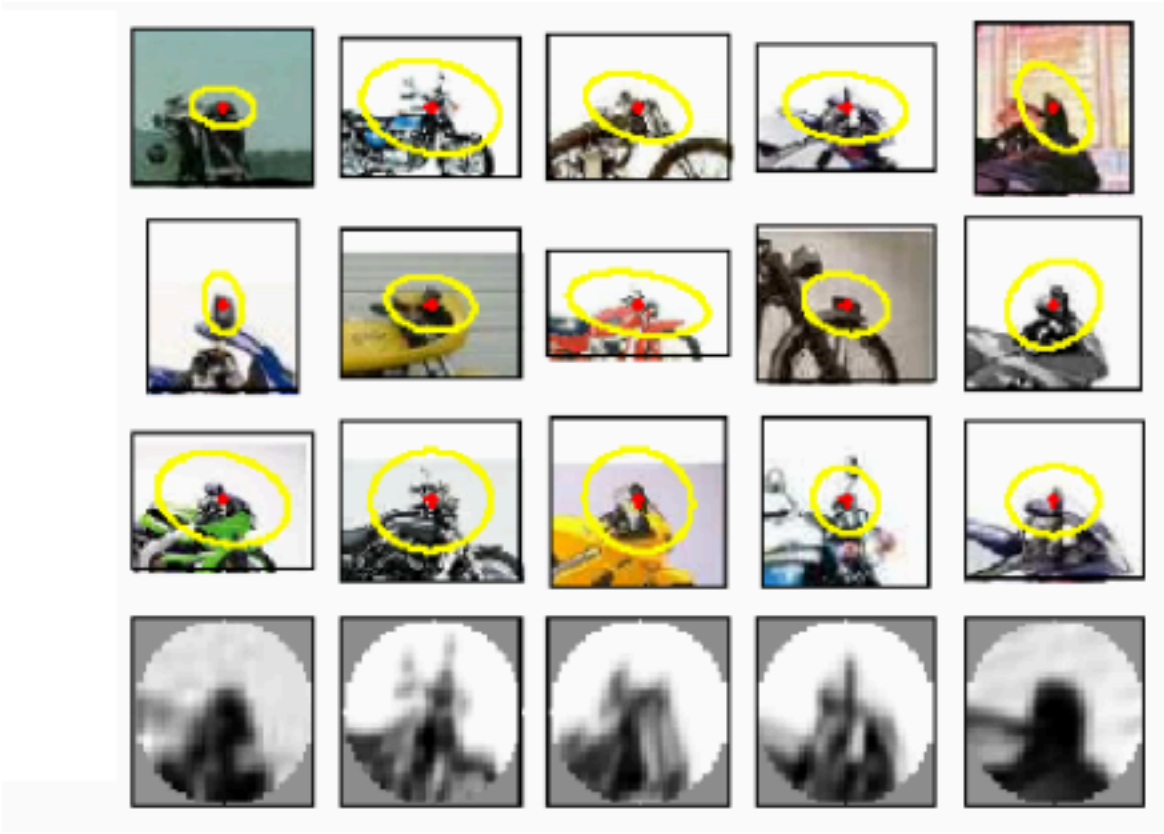
- Feature의 등장

이제 이미지 분류를 꽤 할 수 있어요.
하지만, 사람이 할 수 있는 수준에는
아직 미치지 못해요.

Visual Words 예제 (Feature)

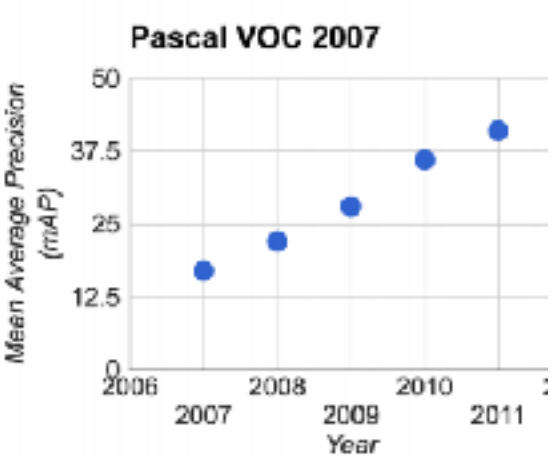


Visual Words 예제 (Feature)



PASCAL Visual Object Challenge
(20 object categories)

[Everingham et al. 2006-2012]



바보야 문제는 알고리즘이 아니라 데이터야

- 페이페이 리 -

여기서 말하는 데이터 문제의 핵심.
1. 데이터의 양이 많아야 한다.
2. 데이터의 레이블이 잘 되어있어야 한다.

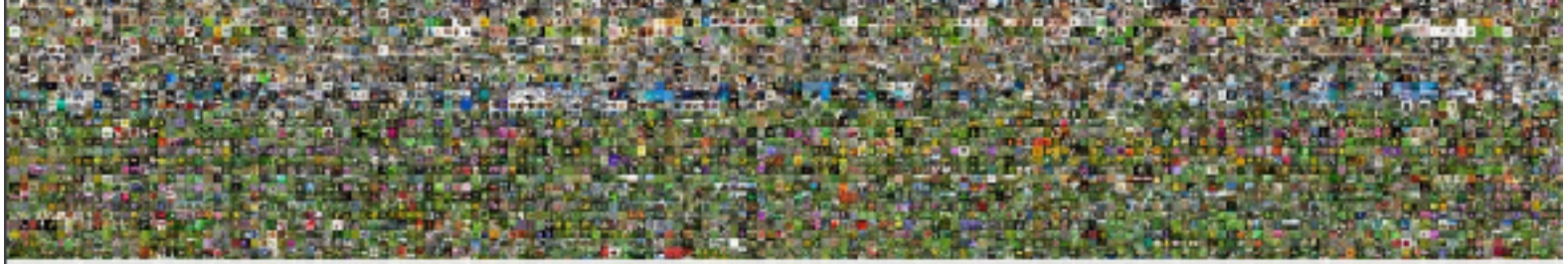
세 살 어린이도 하는데 컴퓨터는 왜 못하는가?

https://youtu.be/WZGty8kjoE0?t=16


TED
아이는 세 살까지 수억장의
현실세계 사진을 보게 됩니다.


IMAGENET


48,940 workers
167 countries

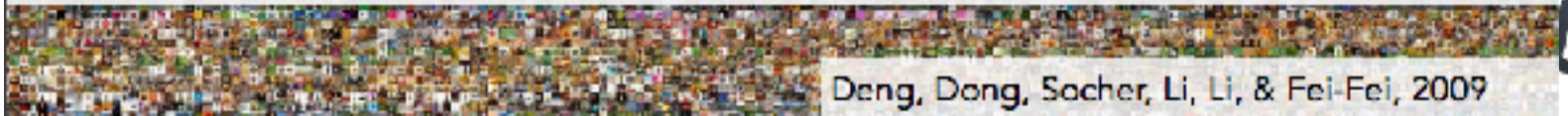


IMAGENET

www.image-net.org

22K categories and **14M** images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities



Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

그 분이 입장하였습니다.

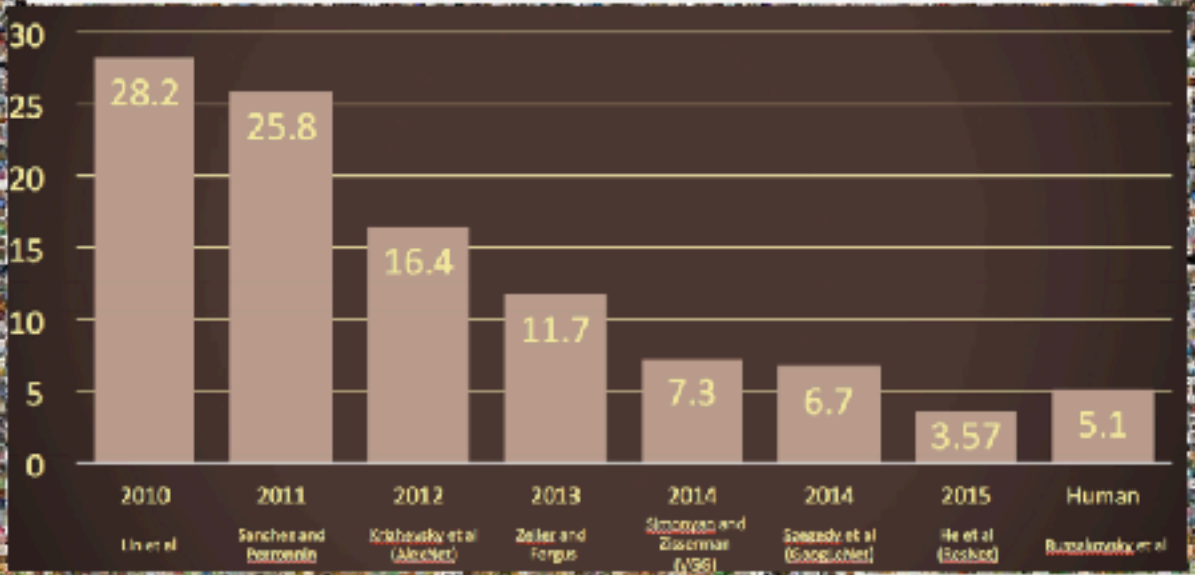
Deep Learning and CNN

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

1,431,167 images



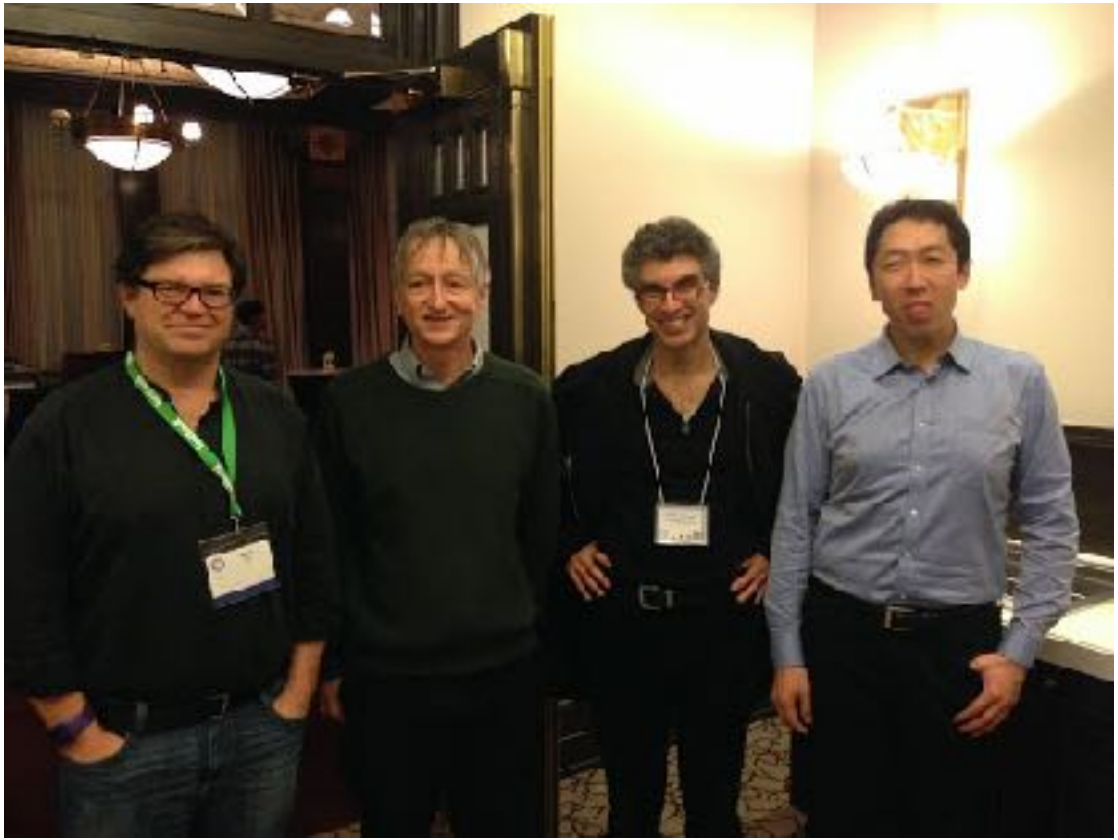
Russakovsky et al. arXiv, 2014

컴퓨터가
어디까지 볼 수 있게 되었나
확인해 보자.

https://youtu.be/WZGty8kjoE0?t=14m13s

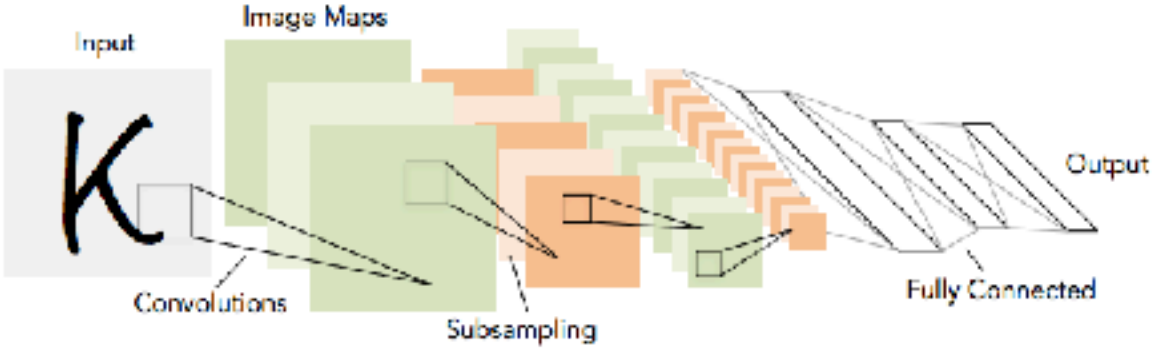
Convolutional Neural Network

CNN은 대체 어떤 마법인가?



Left to right: Yann LeCun, Geoff Hinton, Yoshua Bengio, Andrew Ng at NIPS 2014 (from Andrew Ng's Facebook page).

1998
LeCun et al.



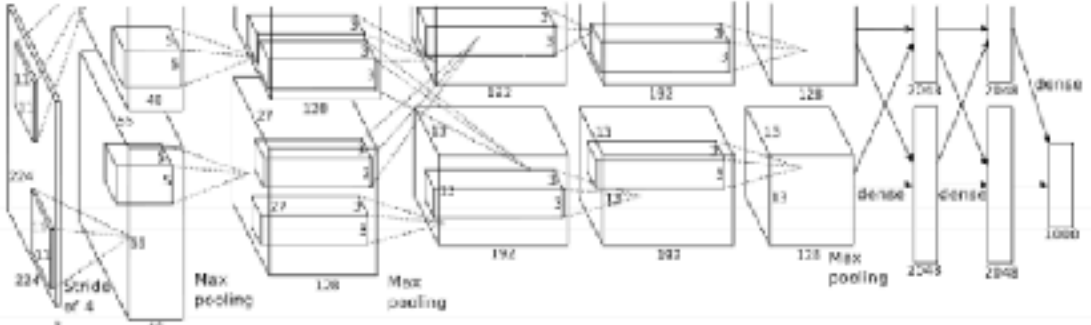
# of transistors

10^6

# of pixels used in training

10^7 NIST

2012
Krizhevsky et al.



# of transistors

10^9

GPUs

# of pixels used in training

10^{14} IMAGENET

Hubel & Wiesel, 1959

Simple cells:
Response to light orientation

Complex cells:
Response to light orientation and movement

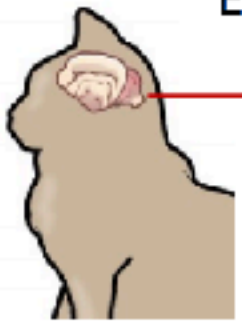
Hypercomplex cells:
response to movement with
nd

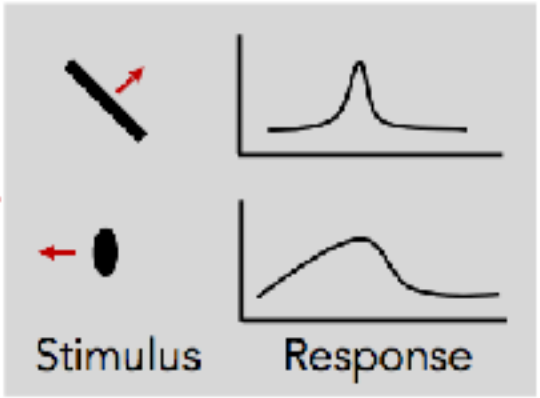
No response

Response (end point)

Stimulus

Electrical signal from brain





1.조각을 본다

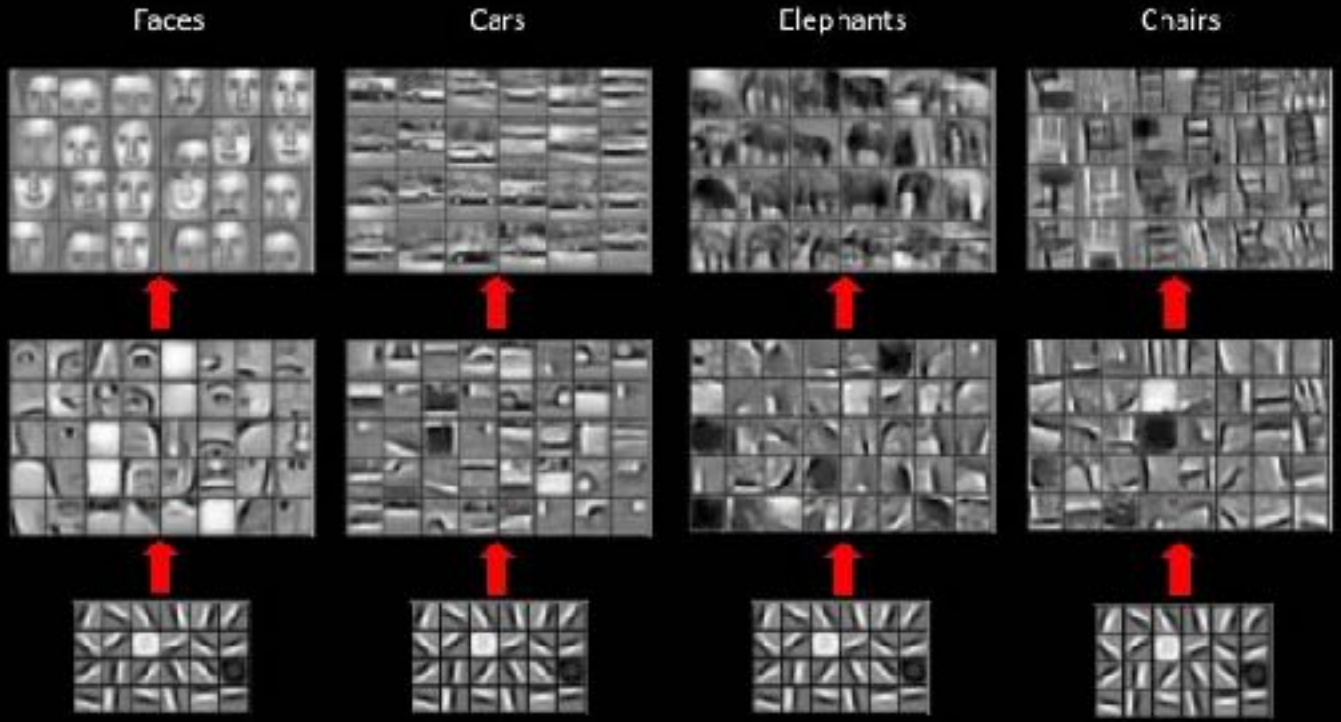
2.각 조각이 조합된 패턴을 본다

3.점점 더 복잡한 조합의 패턴을 본다.

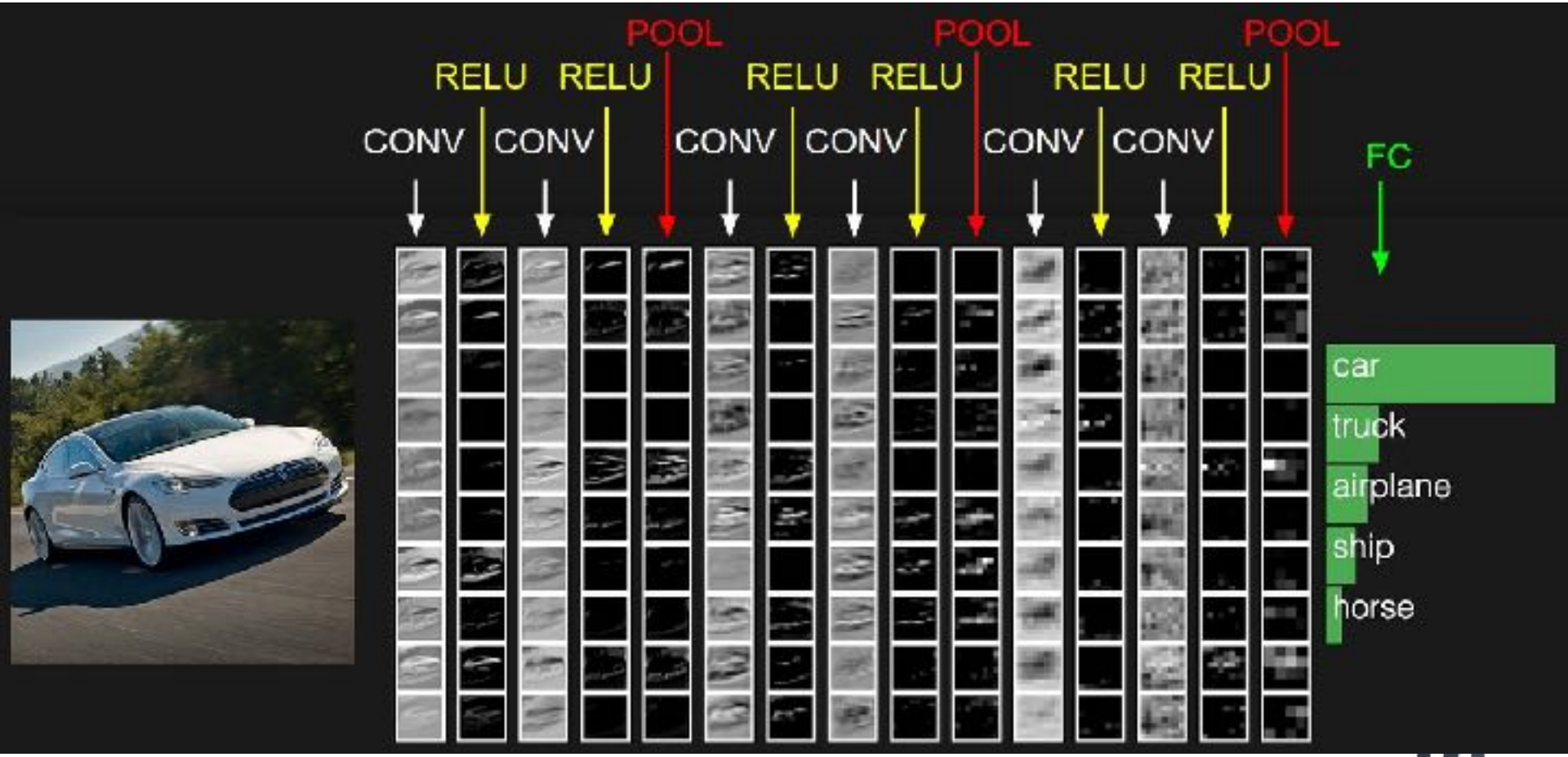
4.반응하는 여러 패턴의 조합을 가지고 이미지를 인식한다.

Learning of object parts

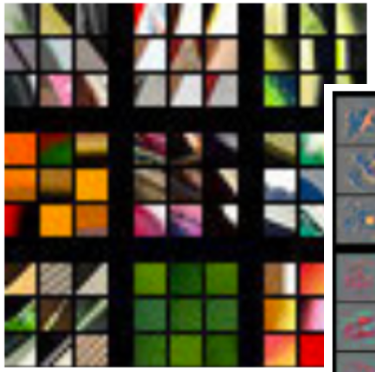
Examples of learned object parts from object categories

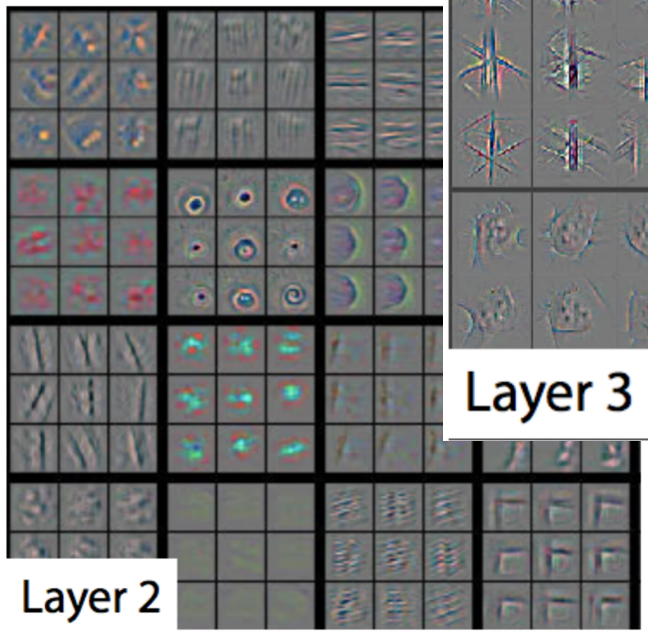


“필터를 CNN이 스스로 학습한다.”

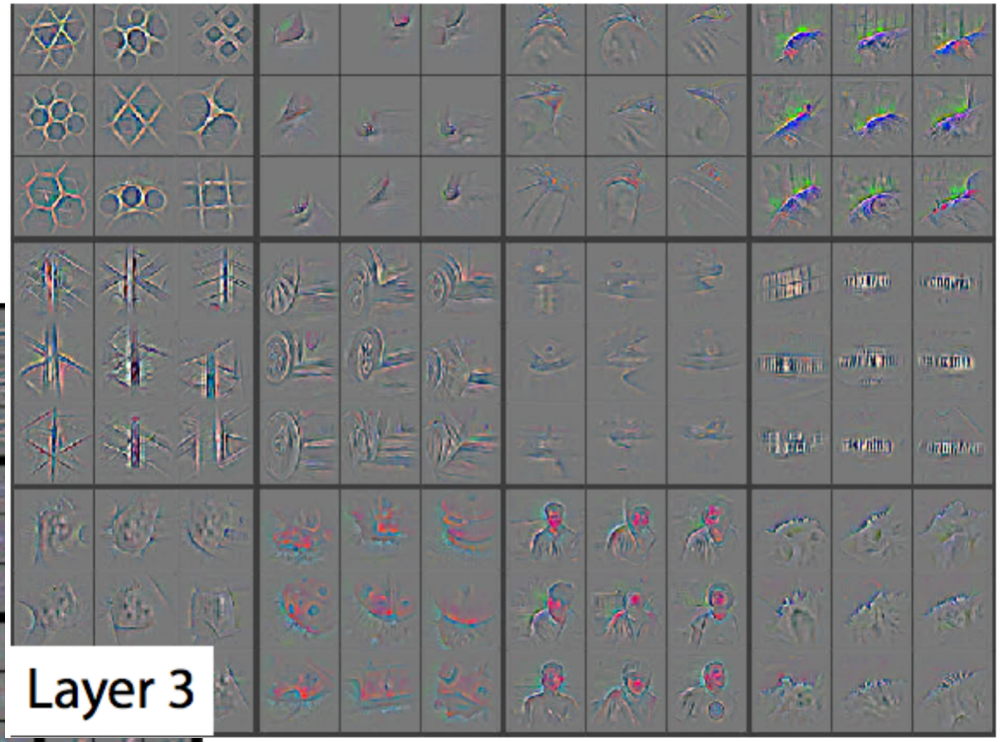

POOL
POOL
POOL
RELU
RELU
RELU
RELU
RELU
RELU
CONV
CONV
CONV
CONV
CONV
CONV
FC
car
truck
airplane
ship
horse

Layer 1





Layer 2



Layer 3

https://arxiv.org/pdf/1311.2901.pdf

자세한 건 다음 시간에...