



58.206.101.71:8266      User: cvpr2019      Pw: xjtucvpr

# 计算机视觉与模式识别

**Computer Vision and Pattern Recognition**

---

## - Introduction

人工智能与机器人研究所

Institute of Artificial Intelligence and Robotics

袁泽剑

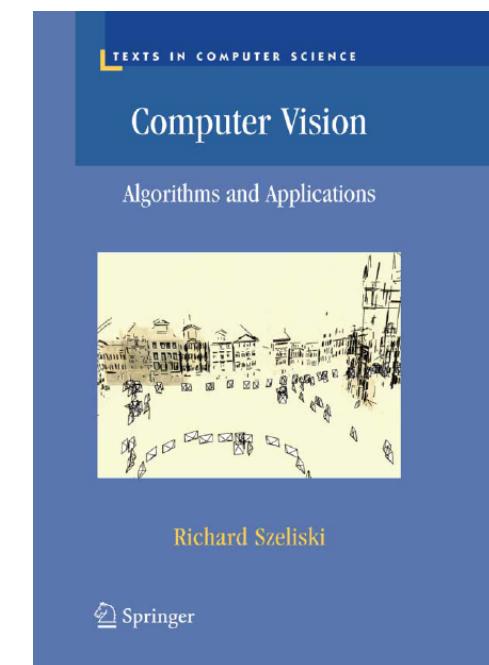
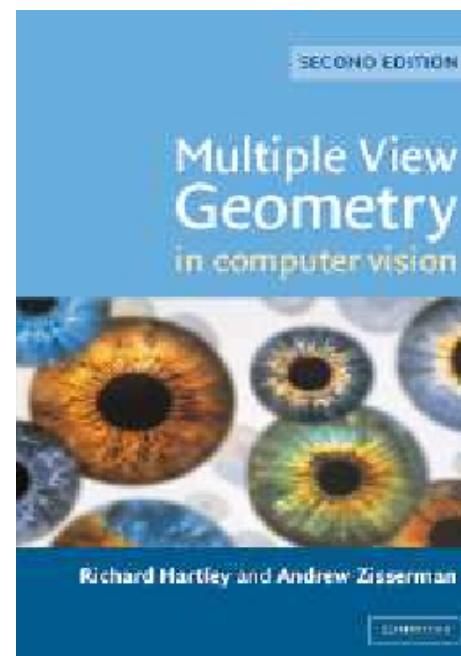
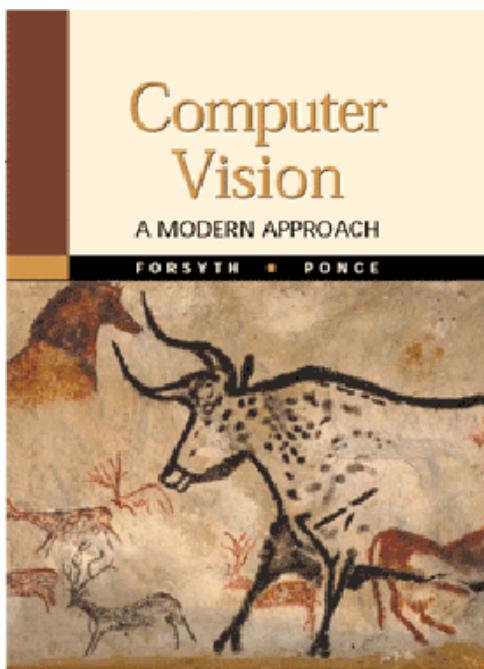
Email: [yuan.ze.jian@xjtu.edu.cn](mailto:yuan.ze.jian@xjtu.edu.cn)

科学馆102



# Textbooks

- D. Forsyth, J. Ponce, **Computer Vision – A Modern Approach**, Prentice Hall, 2002
- R. Hartley, A. Zisserman, **Multiple View Geometry in Computer Vision**, 2nd Ed., Cambridge Univ. Press, 2004
- Richard Szeliski, **Computer Vision: Algorithms and Applications**, Springer, 2011.





# 提纲

---

- 计算机视觉简介
- 课程的主要内容
- 国际期刊与会议
- 主要的参考文献



# 为什么要研究计算机视觉?

## ■ 图像传感器 (Camera) 无所不在(**ubiquitous**)





# 为什么要研究计算机视觉?

## ■ 图像与视频数据无所不在



Personal photo albums



Movies, news, sports



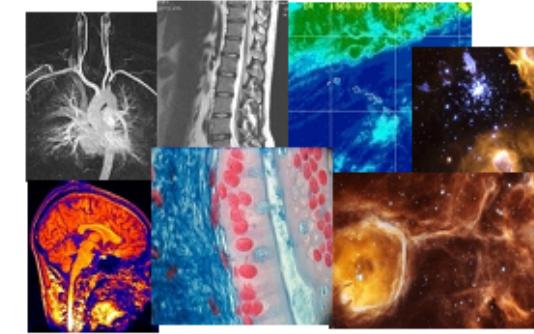
Internet services



Surveillance and security



Mobile and consumer  
applications



Medical and scientific images

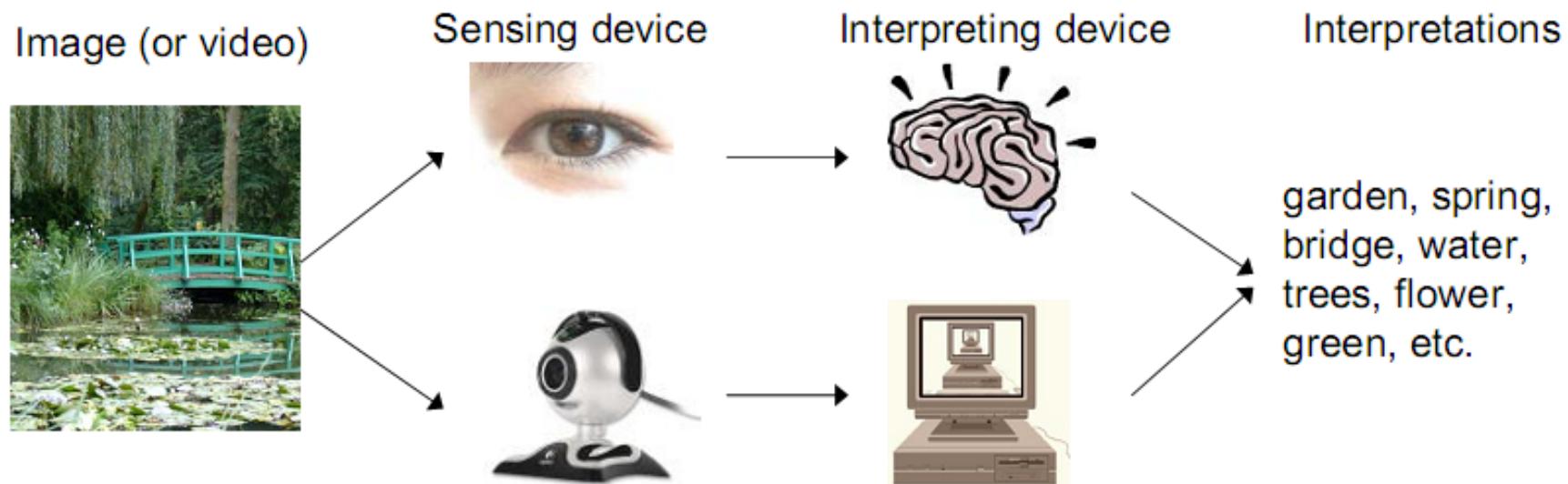
如何实时处理与分析，如何从海量视觉数据中提取有用信息，如何组织、管理与利用，是信息技术面临的主要瓶颈之一。  
Institute of Artificial Intelligence and Robotics -Yuan Zejian



# 什么是计算机视觉？

## What is (computer) vision?

- What does it mean, to see? The plain man's answer (and Aristotle's, too), would be, to know **what is where** by looking[1]. -- David Marr, Vision (1982)
- Computer Vision: The study of how **computers** can be programmed to **extract useful information** about the environment from **optical images** [2]. -- S.E. Palmer, Vision science (1999)



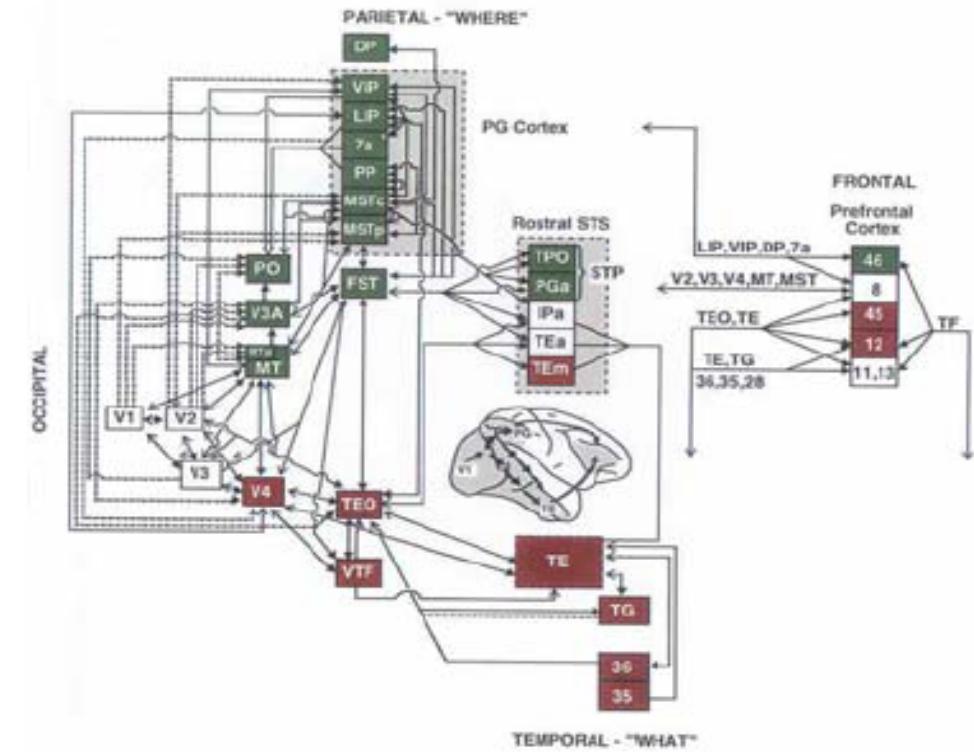
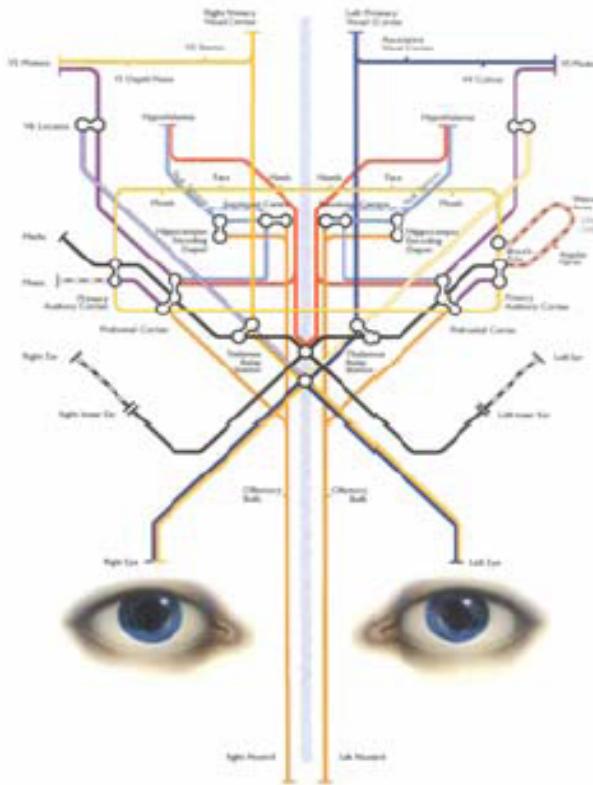
© L. Fei Fei

视觉：去看意味什么？一般人的回答是要知道什么东西在哪什么？  
CV：研究计算机如何通过编程从光学图像中抽取有关环境的有用信息。



# Introduction

- Human hardware & Visual perception



**Visual Perception:** The process of **acquiring knowledge** about environmental objects and events by **extracting information** from the light they emit or reflect [2]. – S.E. Palmer, Vision Science.



# 计算机视觉目的是什么？

---

- 使机器能够理解图像与视频
  - 从视觉数据中计算3D世界的几何属性，或再现3D场景  
(光学测量、3D重建、可视化、增强现实)
  - 建立表示方式与计算方法（算法），使机器能够识别目标、人、物、场景与行为等  
(感知与解释)。



# 计算机视觉到达什么水平？

- Can computers match human perception?
  - computer vision is still no match for human perception
  - but catching up, particularly in certain areas

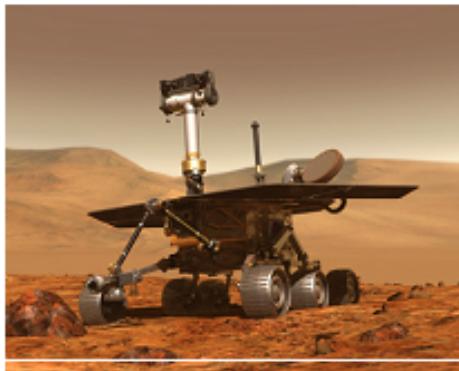


**Smart Erase Microsoft Digital Image Pro:**



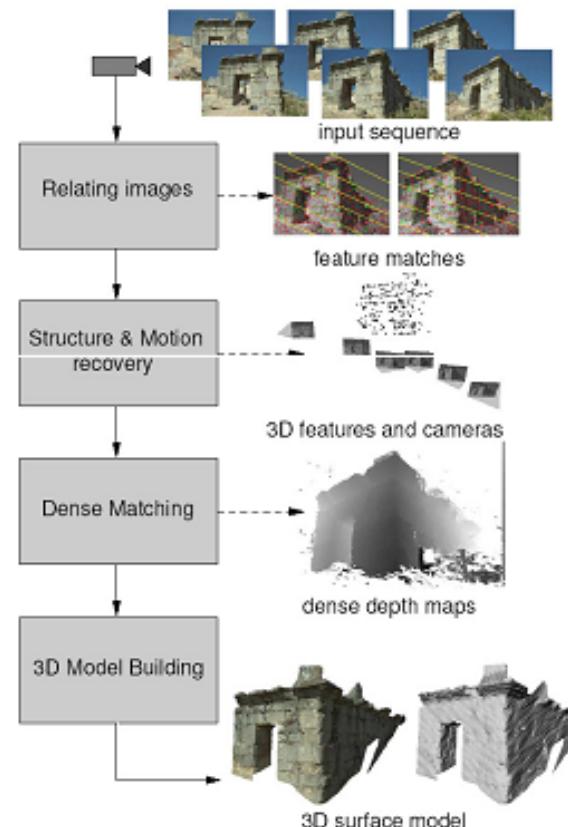
# 视觉重建与可视化

## Real-time stereo



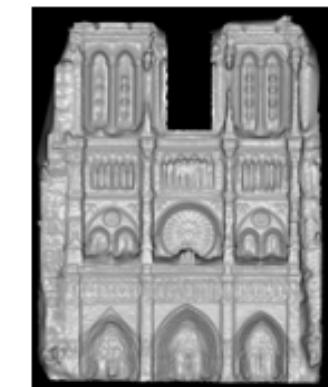
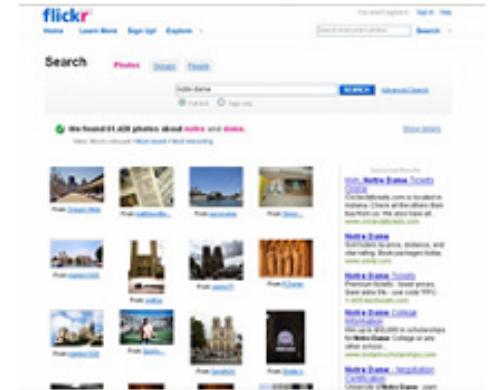
Pollefeys et al.

## Structure from motion



Pollefeys et al.

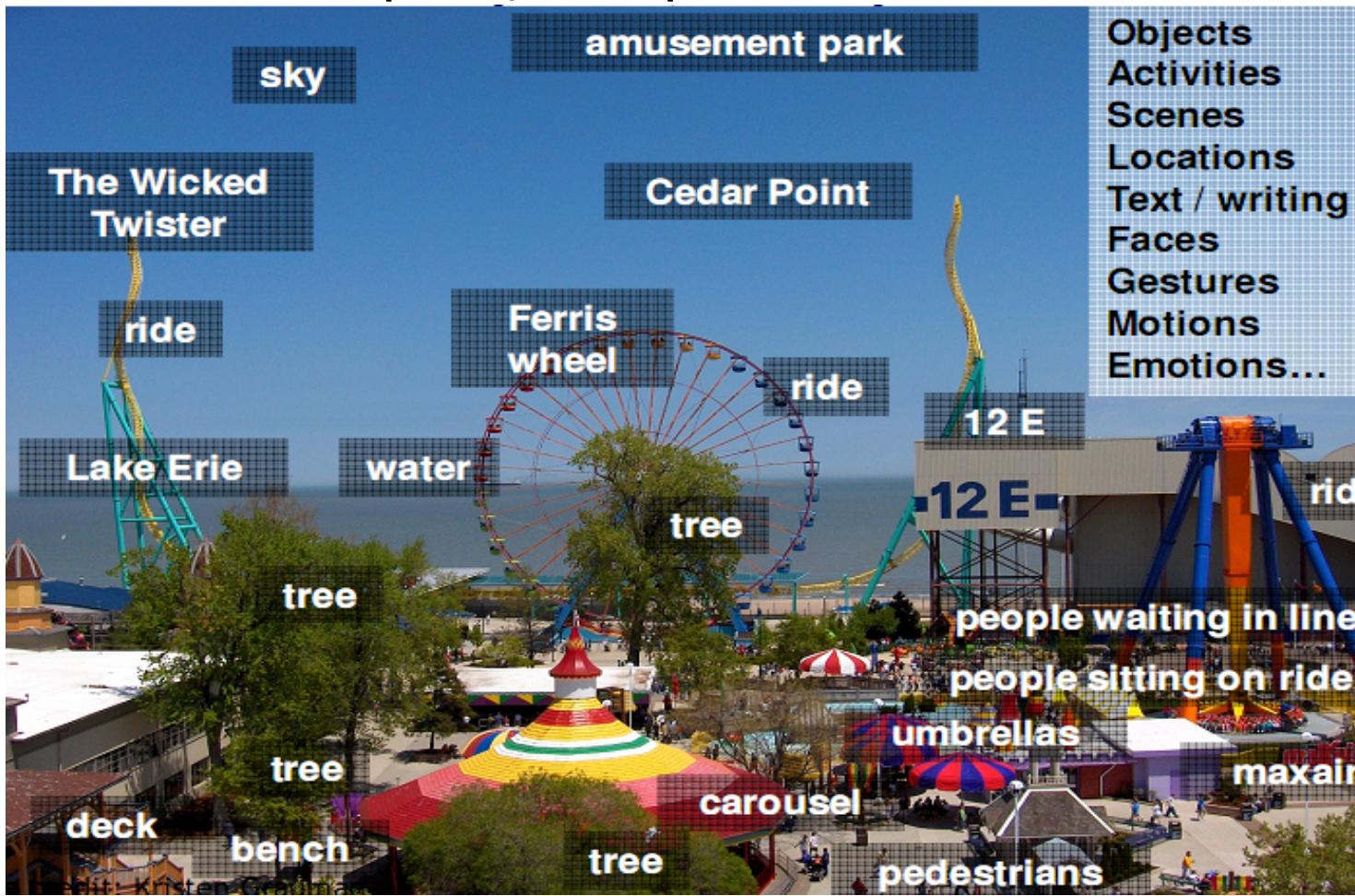
## Multi-view stereo for community photo collections



Goesele et al.



- Vision for Perception, Interpretation



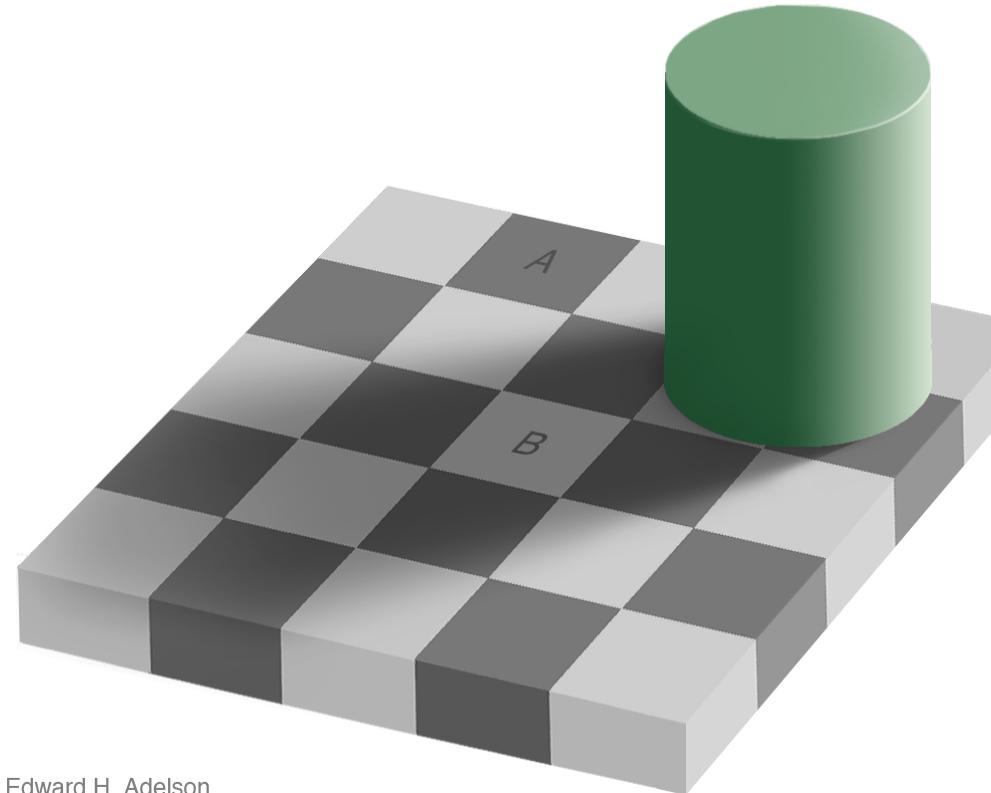
与人类的视觉感知与认知能力相比，还相差甚远



# Introduction

- **Reality and Illusion / Measurement vs. Perception**

Brightness



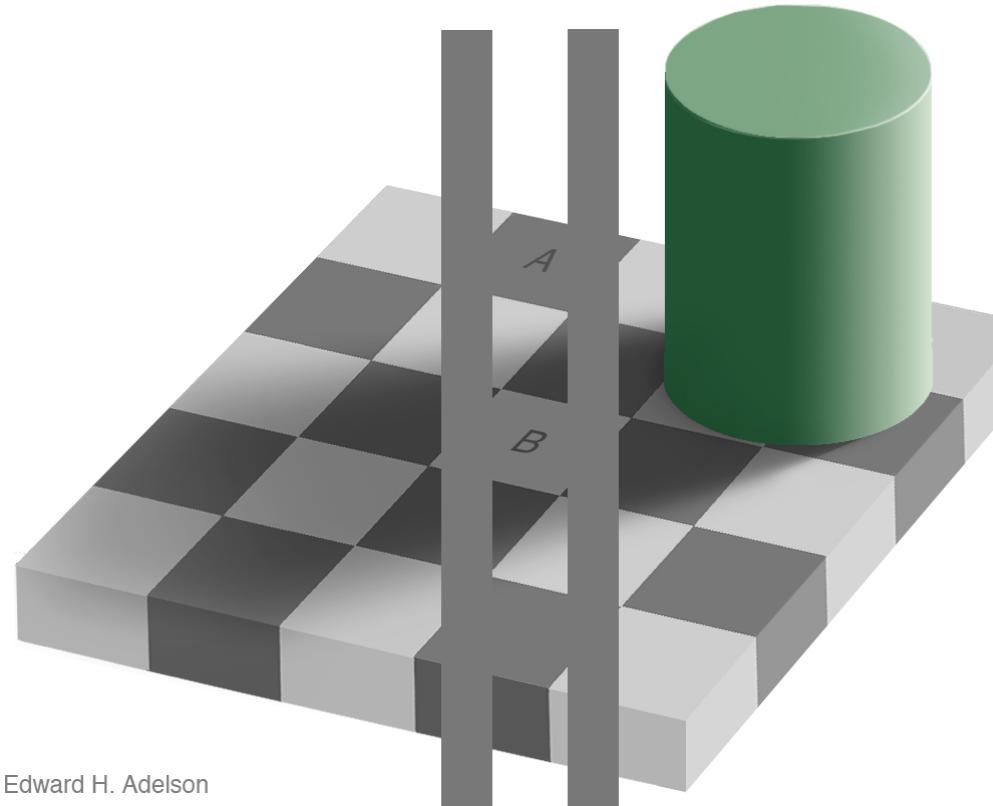
Edward H. Adelson



# Introduction

- **Reality and Illusion / Measurement vs. Perception**

Brightness

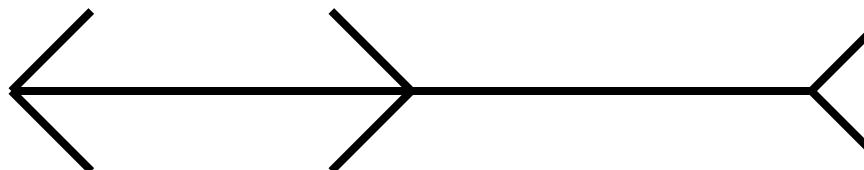


**Proof!**



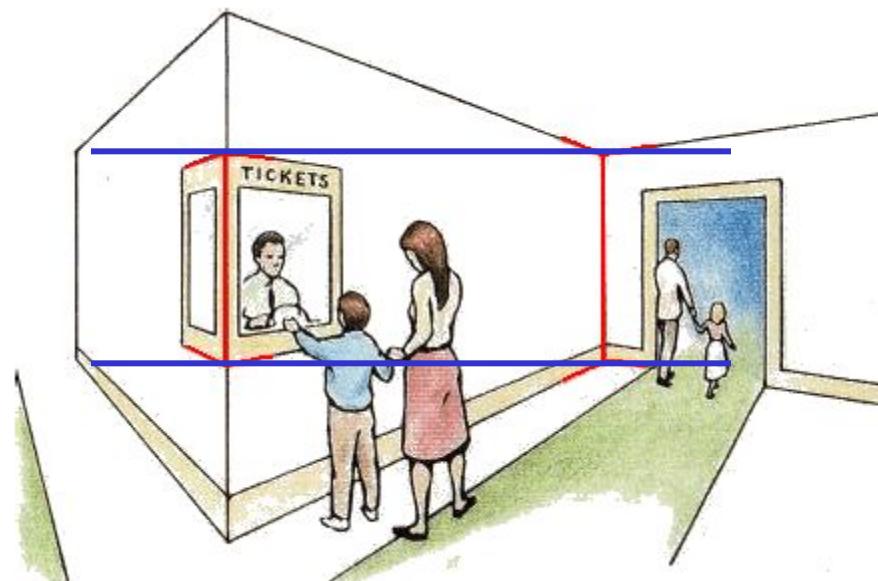
# Introduction

- **Reality and Illusion / Measurement vs. Perception**



Müller-Lyer Illusion

Lengths



[http://www.michaelbach.de/ot/sze\\_muelue/index.html](http://www.michaelbach.de/ot/sze_muelue/index.html)



# Introduction

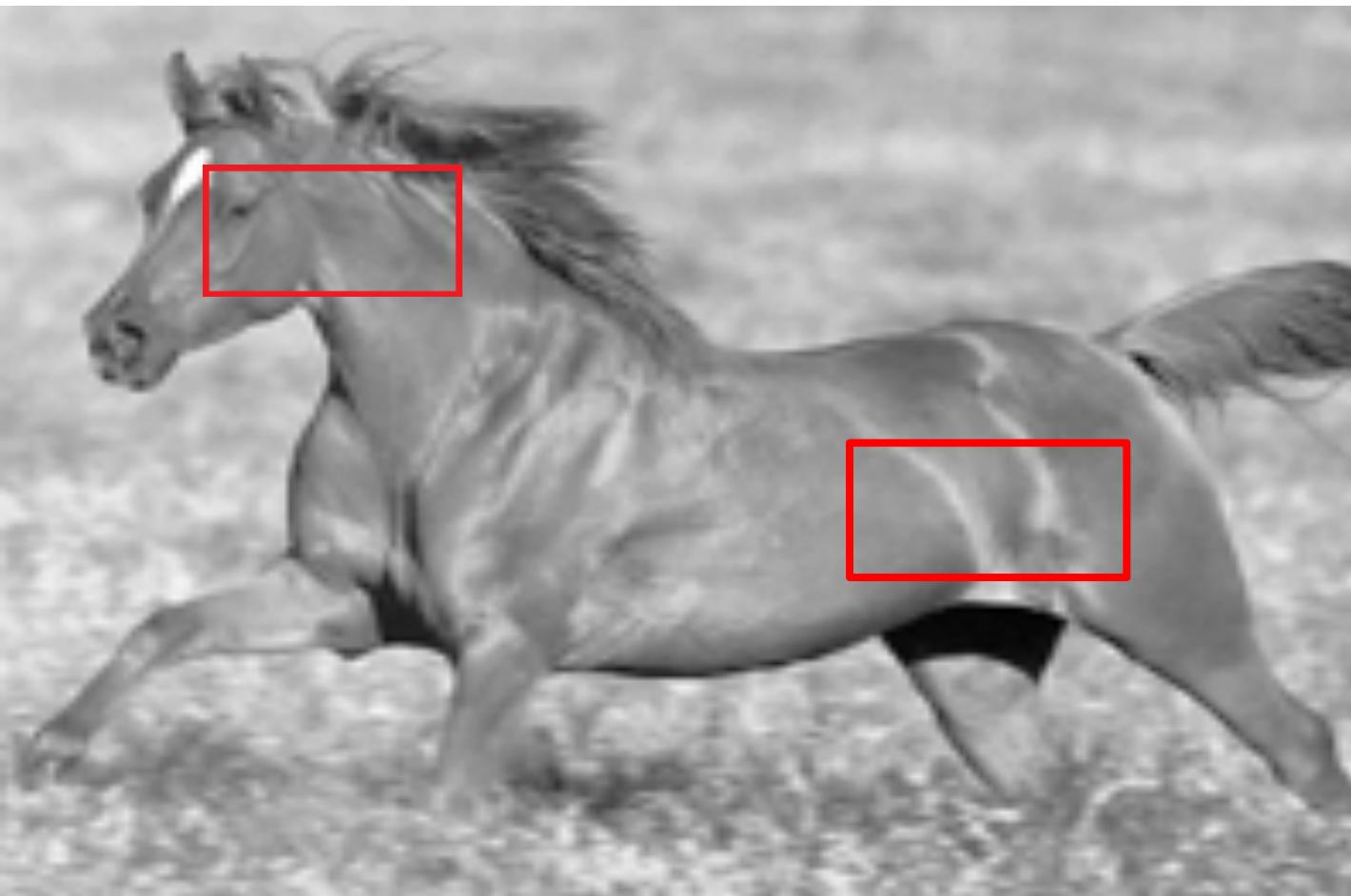
- **Visual Perception needs global information propagation (local & global)**





# Introduction

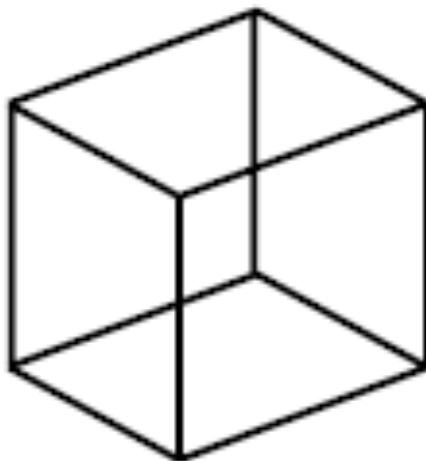
- **Visual Perception needs global information propagation**





# Introduction

- Perception could be ambiguous



Necker cube



Vase vs. face



bikini vs. martini

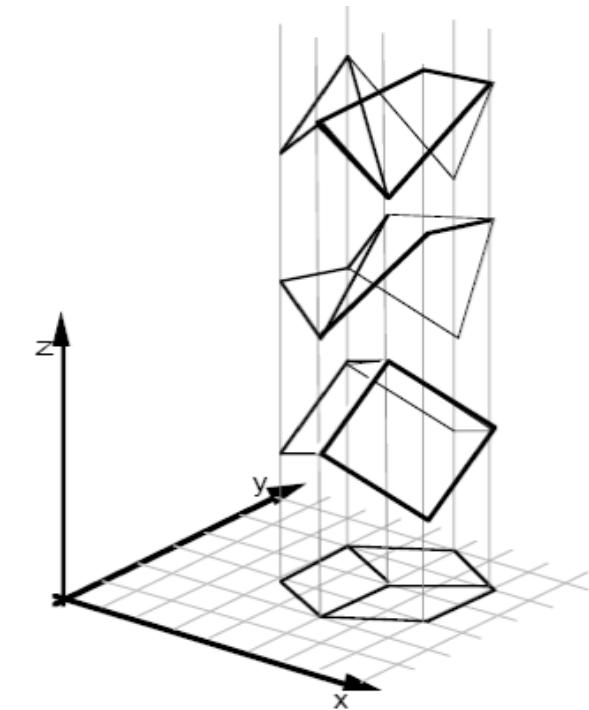
Single image can give rise to two or more distinct perceptions.

- Mutually exclusive
- Multistable perceptions



# Introduction

- Vision is ill-posed
- > EXAMPLE:
- > Recovering 3D geometry from **single** 2D projection
- > Infinite number of possible solutions!
- > Concept (Visual **categories**) & Image (Visual **instances**)



from [Sinha and Adelson 1993]



# Introduction

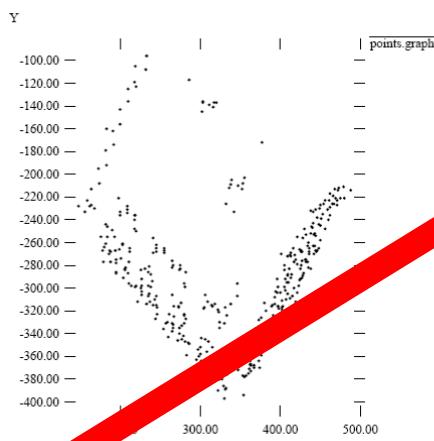
## ▪ Vision as Measurement Device



Real-time stereo on Mars



Physics-based Vision



Structure from Motion



Virtualized Reality



# Introduction

---

- Small Reason:
  - For measurement, other sensors are often better (in DARPA Grand Challenge, vision was barely used!)
- Big Reason:

The goals of computer vision (**What + Where**) are in terms of what humans care about.

计算机视觉的目标重点在人类所关心的方面



# Introduction

## What do humans care about?



slide by Fei Fei, Fergus & Torralba



# Introduction

## Verification: is that a bus? (1-1)





# Introduction

**Detection:** are there cars?





# Introduction

**Identification: is that a picture of Mao? (1- n)**

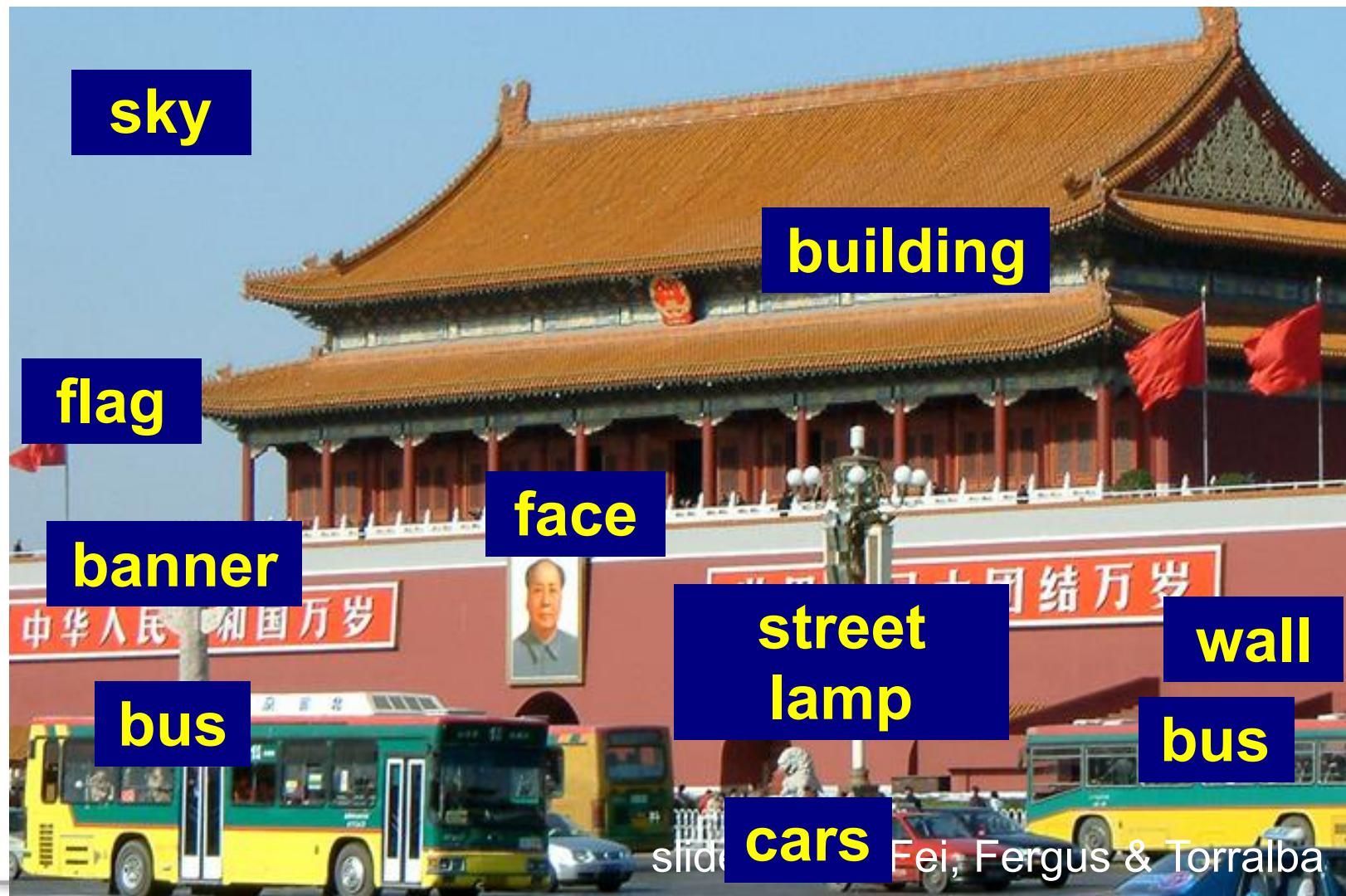


slide by Fei Fei, Fergus & Torralba



# Introduction

## Object categorization





# Introduction

## Scene and context categorization

- **outdoor**
- **city**
- **traffic**
- ...



slide by Fei Fei, Fergus & Torralba



# Introduction

## Rough 3D layout, depth ordering

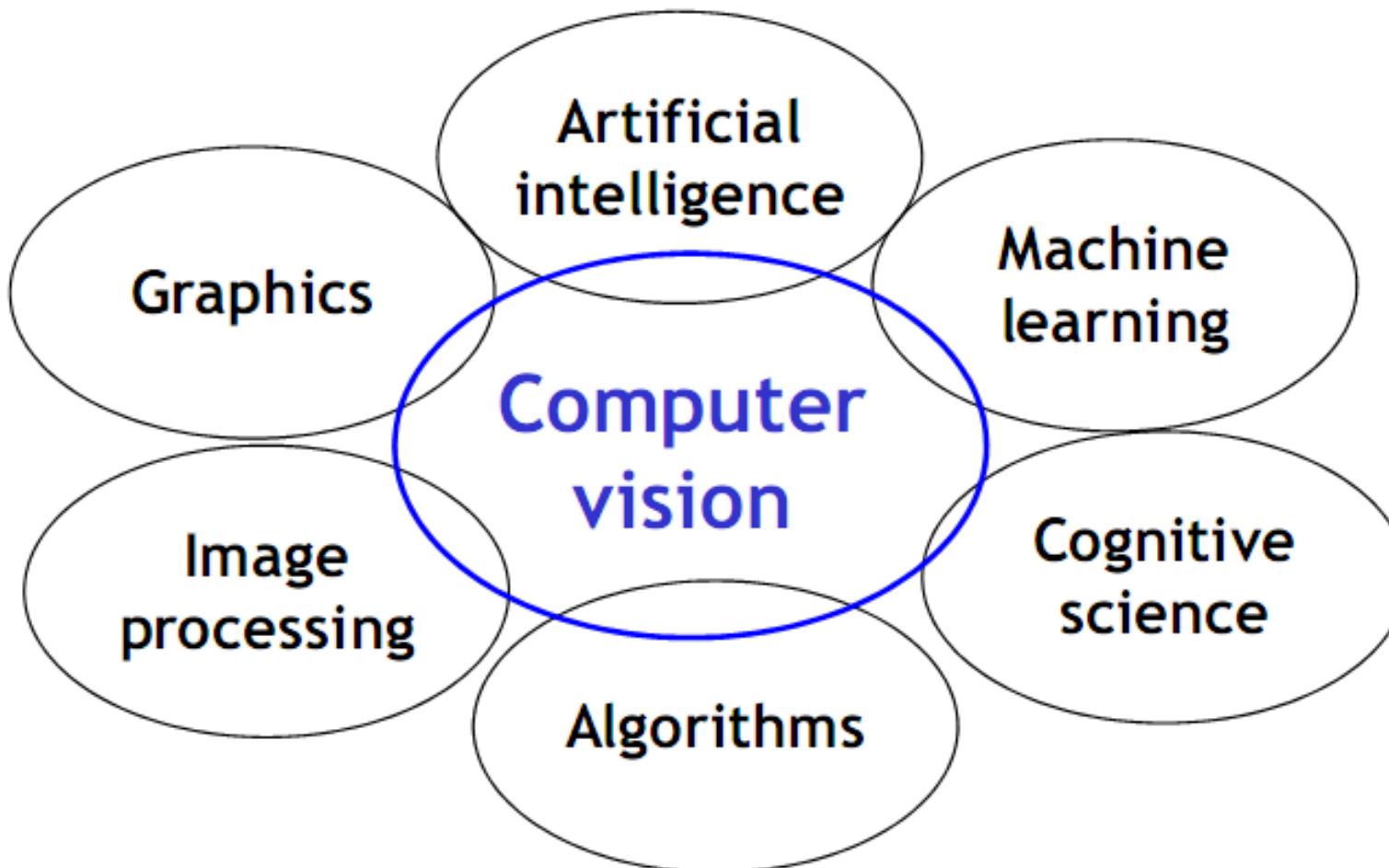


slide by Fei Fei, Fergus & Torralba



# Introduction

## Related Disciplines





# Directions to Computer Vision

---

- Science
  - ✓ Foundations of perception. How do WE see?
- Engineering
  - ✓ How do we build systems that perceive the world?
- Many applications
  - ✓ Searching , surveillance, Medical imaging, entertainment, graphics, ...



# Applications: Faces and Digital Cameras

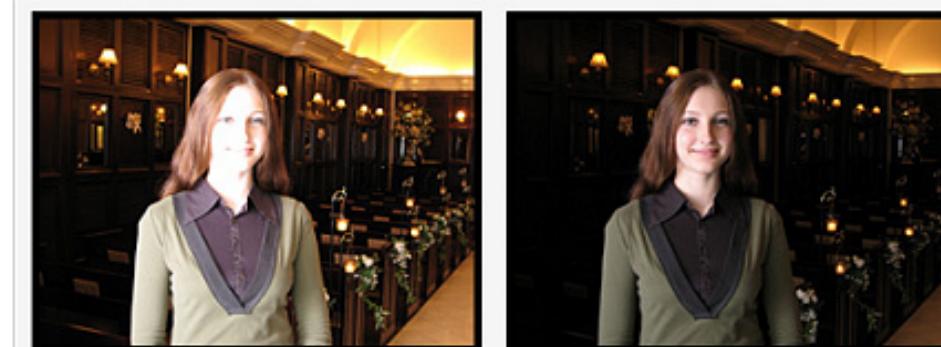


Setting camera focus  
via face detection



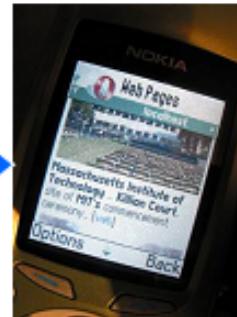
Camera waits for everyone to  
smile to take a photo [Canon]

Automatic lighting  
correction based  
on face detection





# Applications: Vision for Mobile Phones



Situated search  
Yeh et al., MIT



MSR Lincoln



kooaba

kooaba

Casino Royale

- [Cinemax: Reviews, Trailer](#)
- [FilmBlog.ch](#)
- [Amazon Mobile](#)
- [Ebay Mobile](#)
- [MSN Mobile Movies](#)
- [Zinnova Mobile](#)

[Cell Kitag for Ticket](#)

[Tell a friend \(by SMS\)](#)

[Home](#)

Search for another movie title on our movie portal:

Download the kooaba client for even easier mobile search!

通过图像识别搜索相关网络信息。



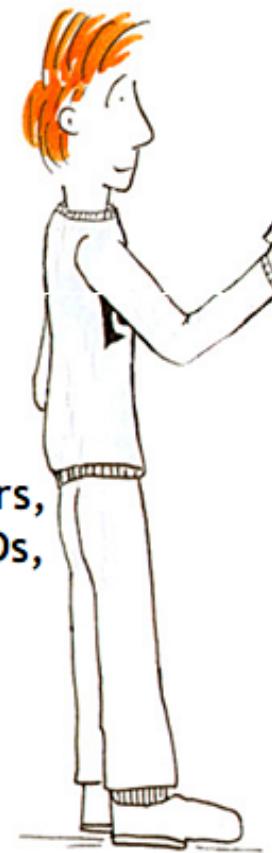
# Demo: Movie Poster Recognition

- Commercial services coming out...

~7 million images indexed (movie posters, book covers, CDs, DVDs, video games)

Query-by-image from mobile phone available in Germany and Switzerland

MOBILE IMAGE RECOGNITION?  
TRY IT OUT NOW!!!



**kooaba**

AN INSTANT CLASSIC!  
TOMMY LEE JONES JAVIER BARDEM JOSH BROLIN  
NO COUNTRY FOR OLD MEN  
A COEN BROTHERS FILM  
THERE ARE NO CLEAN GETAWAYS

Show another poster

Movie data provided by:

1. POINT YOUR MOBILE PHONE CAMERA TO THE MOVIE POSTER.

2. SNAP A PICTURE AND SEND IT:

IN SWITZERLAND:  
MMS TO 5555 (OR  
079 394 57 00  
FOR ORANGE CUSTOMERS)

IN GERMANY:  
MMS TO 84000

EVERYWHERE:  
EMAIL TO  
M@KOOABA.COM

3. FIND ALL RELEVANT INFORMATION ABOUT THE MOVIE ON YOUR MOBILE PHONE

Source: <http://www.kooaba.com>



# Applications: Vision-based Interfaces



**Games**  
**(Sony EyeToy)**



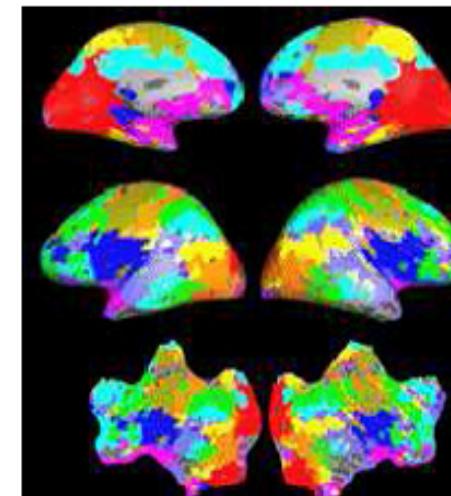
**Assistive technology systems**  
**Camera Mouse**  
**Boston College**



# Applications: Medical & Neuroimaging



Image guided surgery  
MIT AI Vision Group



fMRI data  
Golland et al.



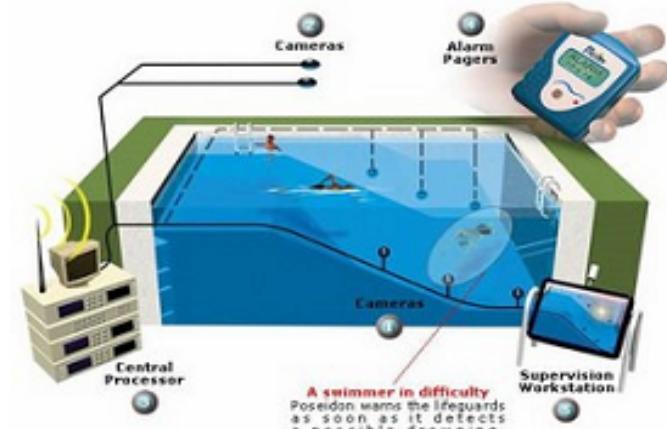
# Applications: Safety & Security



Autonomous robots



Driver assistance



Monitoring pools  
(Poseidon)



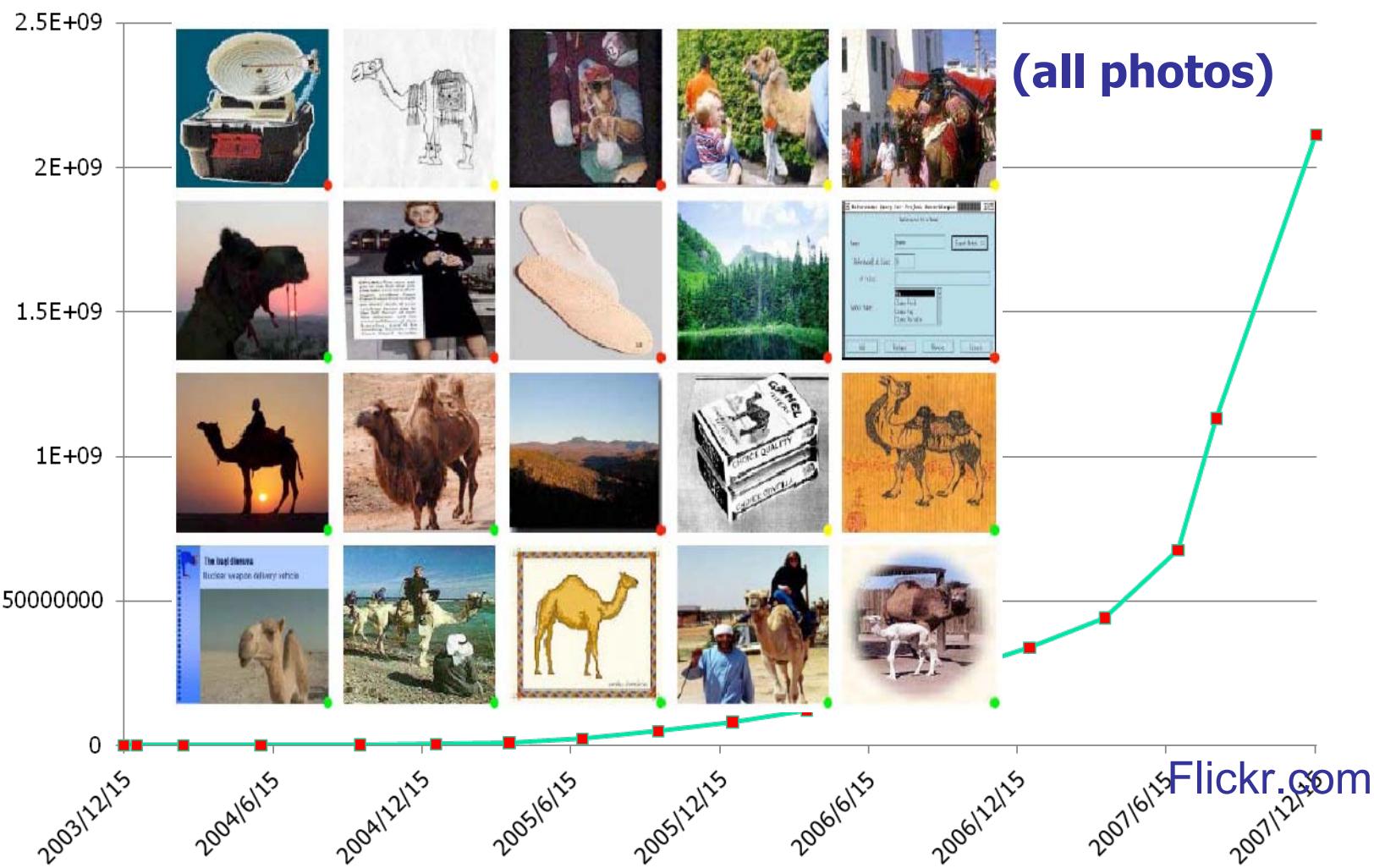
Pedestrian detection  
[MERL, Viola et al.]



Surveillance

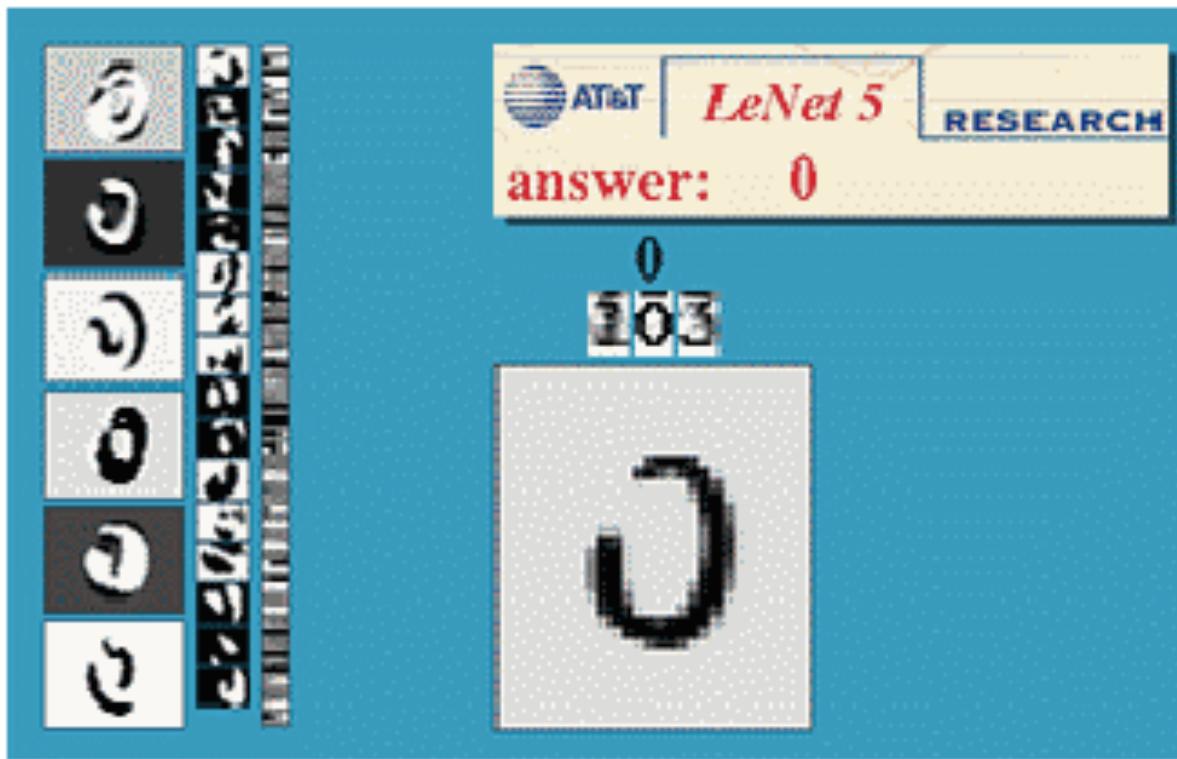


# Applications: Searching the web





## Applications: Document Analysis



Digit recognition, AT&T labs

<http://www.research.att.com/~yann/>



# Applications: Graphics and Art





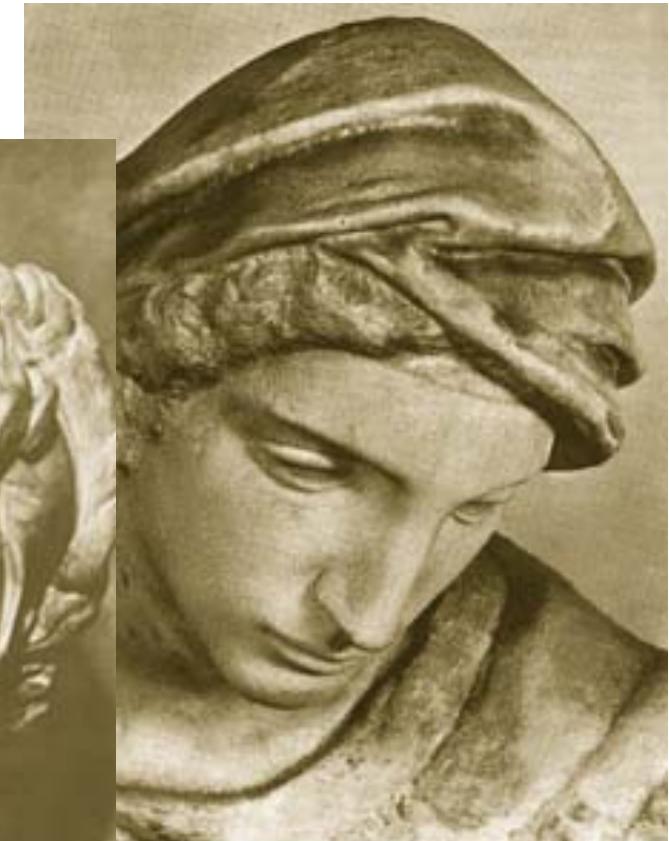
# Applications: Toys and robots

## Robotics Vision





# Challenges 1: view point variation



Michelangelo 1475-1564



## Challenges 2: illumination



slide credit: S. Ullman



# Challenges 3: occlusion





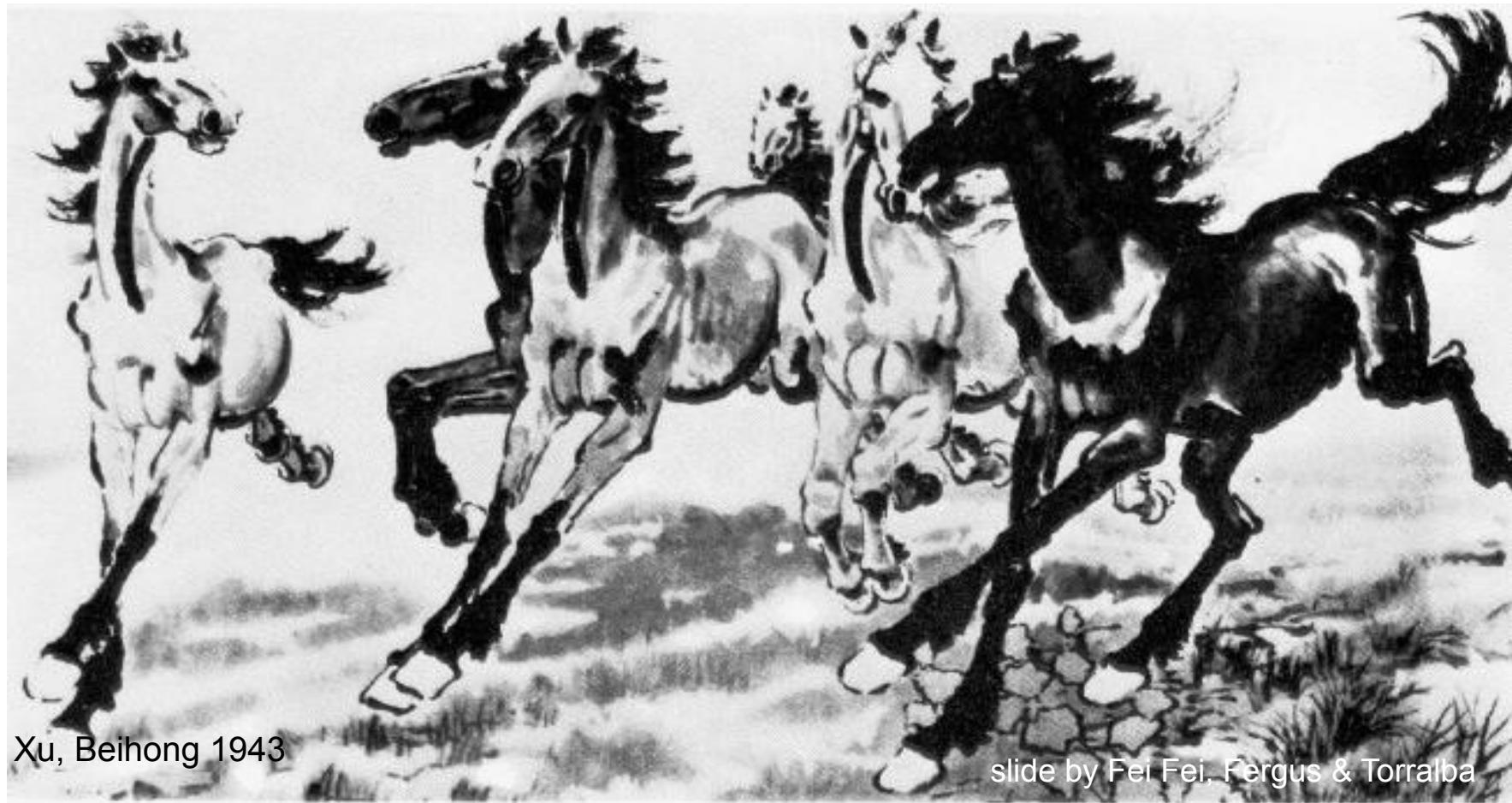
# Challenges 4: scale



slide by Fei Fei, Fergus & Torralba



# Challenges 5: deformation



Xu, Beihong 1943

slide by Fei Fei, Fergus & Torralba



# Challenges 6: background clutter





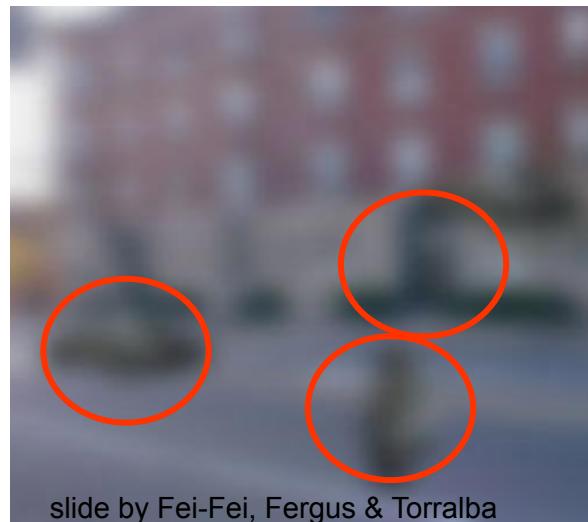
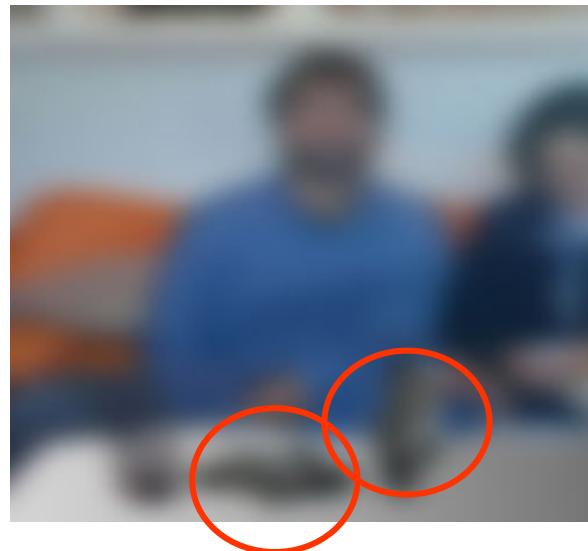
# Challenges 7: object intra-class variation



slide by Fei-Fei, Fergus & Torralba



# Challenges 8: local ambiguity



slide by Fei-Fei, Fergus & Torralba



# Challenges 9: the world behind the image



When people see an image, they see not a plane with patterns of color and texture, but the world behind the image.



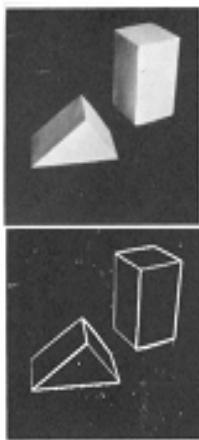
## Challenges 10: Complexity (unstructured data)

---

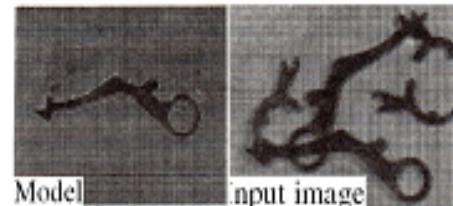
- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images indexed by Google Image Search
- 18 billion+ prints produced from digital camera images in 2004
- 295.5 million camera phones sold in 2005
- **About half of the cerebral cortex** in primates is devoted to processing visual information [Felleman and van Essen 1991]



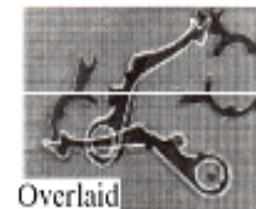
# Exciting progress



.....



.....



....



....



....

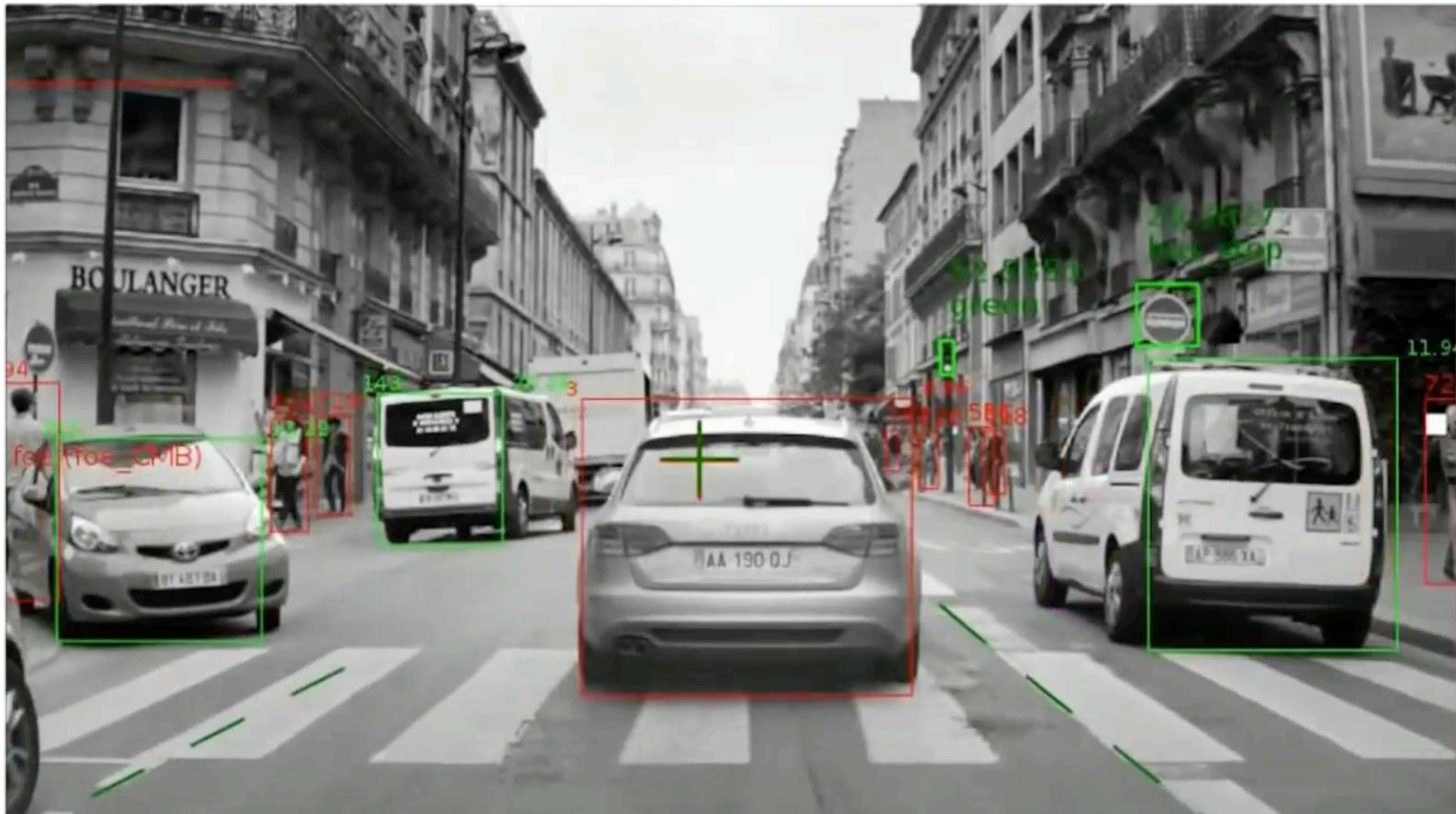


....



# Exciting progress

- Computer Vision in realistic scenarios is becoming feasible!





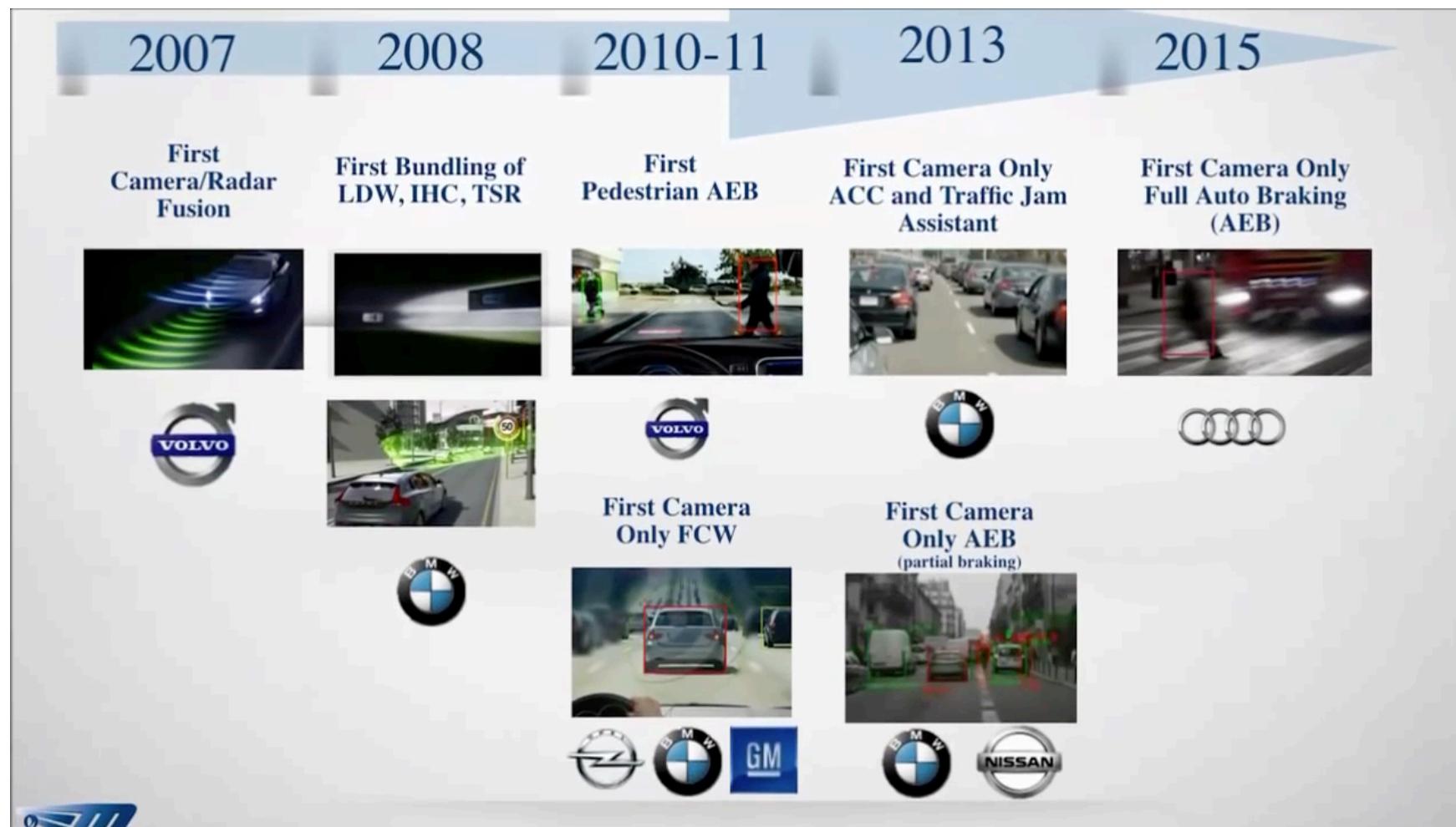
# Exciting progress





# Exciting progress

- Computer Vision in realistic scenarios is becoming feasible!



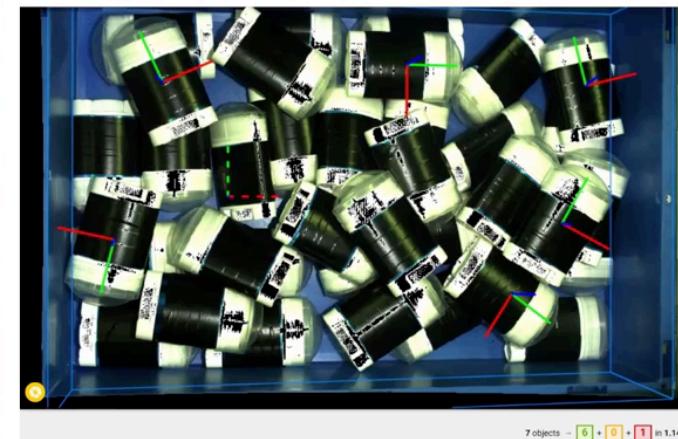


# Exciting progress

## Fast bin picking



Picks      Time  
**03**      **00:05**



**Pickit**  
ROBOT VISION MADE EASY



# Exciting progress

## Automated Shoe Lacing





# Introduction

- Syllabus ( 5 topics )
  - **Physiology and theories of Vision (1)**

- **Camera Model & Camera Calibration (1)**
- **Stereo Vision & stereo matching (1)**
- **Epipolar constraint & E/F estimation (1)**
- **Structure from Motion (SFM) (1)**

Spatial Vision /  
Geometric Vision

- **Visual Motion & Motion Analysis (1)**
- **Optical flow & Optical flow estimation (1)**
- **Object motion & Object tracking (1)**

Motion Vision

- **Image Filtering & CNN (1)**
- **Structure Extraction & Template Matching(1)**
- **Local Features & Features Matching (1)**

Early Vision

- **Perceptual Organisation /Visual Grouping(2)**
- **Graphical model & Inference method (1)**

Mid-level Vision

- **Visual Recognition / Traditional Methods (1)**
- **Visual Recognition / CNN based Methods (1)**

High-level Vision



# 1、Geometric Vision

Image Formation	<b>Image formation; Camera model; Camera parameters; Calibration</b>
Stereo Matching & Reconstruction	<b>Stereo Vision;</b> <b>Triangulation, disparity;</b> <b>Epipolar geometry (Calib.);</b> <b>Essential matrix;</b> <b>Stereopsis &amp; Robust reconstruction</b>
Geometric transformations estimation	<b>Epipolar geometry (Uncalib.);</b> <b>Foundation matrix;</b> <b>Eight-Point Algorithm;</b> <b>RANSAC for robust estimation</b>
Structure from Motion	<b>Structure from motion (SFM);</b> <b>Affine/Euclidean SFM;</b> <b>Projective SFM (Two cameras);</b> <b>Bundle adjustment (Multi-view)</b>



## 2、Motion perception & Tracking

Visual motion	<b>3D motion field;</b> <b>2D image motion &amp; 2D motion field;</b> <b>Motion analysis;</b>
Optical Flow	<b>Apparent motion perception;</b> <b>Optical flow computation methods</b> <b>(Nonparametric &amp; Parametric);</b> <b>KLT feature tracking;</b>
Object Tracing	<b>Tracking with dynamics;</b> <b>Prediction and correction;</b> <b>LDS &amp; Kalman filter/ Particle filter;</b> <b>Object representation; Observation model</b> <b>(visual tracking);</b> <b>Data association;</b>



### 3、Visual features & Matching

Image Filtering	<b>Linear filter; Correlation / convolution; Non-linear filter; Linear scale space ; Filter bank; Multi-scale representation</b>
Edge detection & Structure Extraction	<b>Edge detection; Template matching; Line fitting; Hough Transform; Shape matching</b>
Local Features	<b>Interest points; Harris detector; Scale invariance; Descriptors (SIFT)</b>
Features Matching & Image alignment	<b>Features matching; Image alignment: (2D Affine, 2D Perspective projection-Homography, Global transform constrain ); RANSAC for robust estimation;</b>



## 4、Visual grouping & Recognition

Segmentation (1) (Clustering / Grouping)	<b>Gestalt theory; Feature space; K-means clustering; GMM (EM algorithm); Mean Shift algorithm</b>
Segmentation (2) (Clustering / Grouping)	<b>Graph theoretic segmentation; Spectral clustering (Ncuts); Hierarchical segmentation; Energy minimization methods; (MRF, CRF, Graph cut, BP)</b>
Recognition (1) (Histogram)	<b>Global representation (Histogram); Comparison histogram (Measure); Multidimensional histogram; Object indexing by histogram</b>
Recognition (2) (Subspace)	<b>Subspace representation; Image sparse representation; PCA; FLD/LDA; Robust PCA; NMF; Manifold</b>



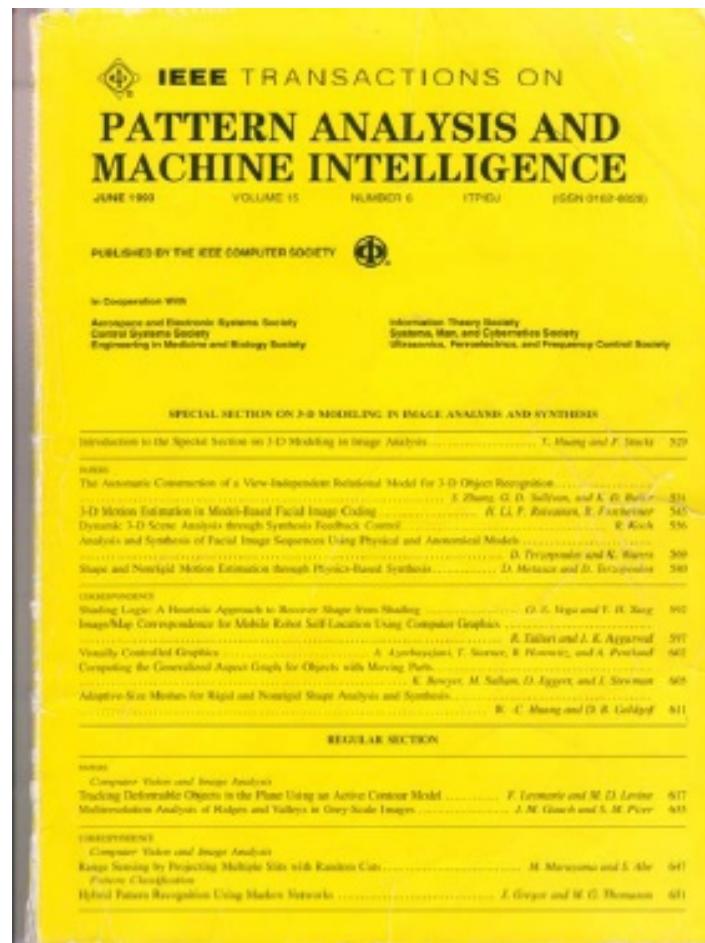
## 4、Visual grouping & Recognition

Recognition (3) (Traditional Methods/ Sliding Window)	<b>Object detection; Identification &amp; categorization; Sliding window (SW) method; Object centered global representation (HOG); Boosting; SVM / LSVM; Beyond Sliding Window</b>
Recognition (4) (Part Based)	<b>Word/Part based method; Bag of words (vocabularies); Pictorial structure model; Constellation model / Star model; Implicit Shape Model (ISM); Deformable Part Model (LSVM), HCRF</b>
Recognition (5) (CNN Based)	<b>DCNN for Image Recognition, Object Detection and Segmentation</b> <a href="http://vision.stanford.edu/teaching/cs231n/syllabus.html">http://vision.stanford.edu/teaching/cs231n/syllabus.html</a>



# Introduction

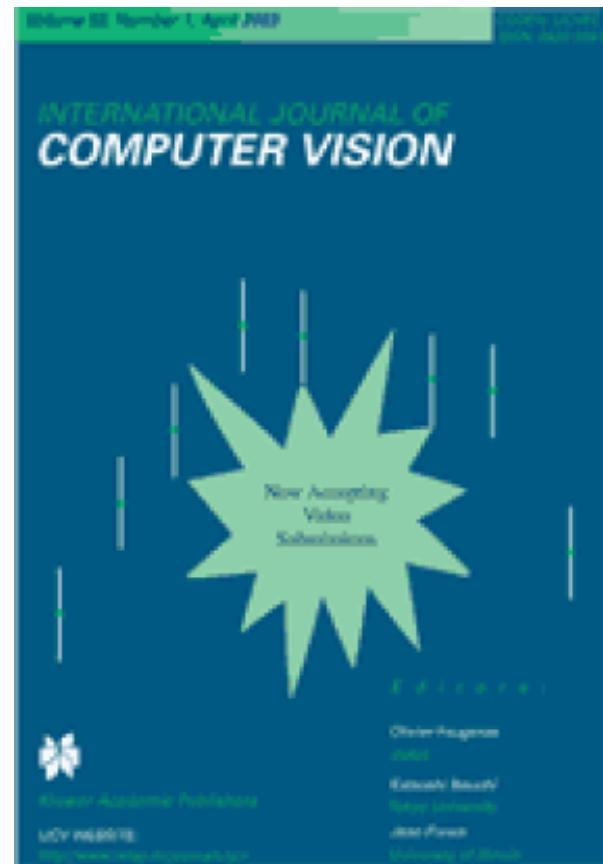
- IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).





# Introduction

- International Journal of Computer Vision (IJCV)

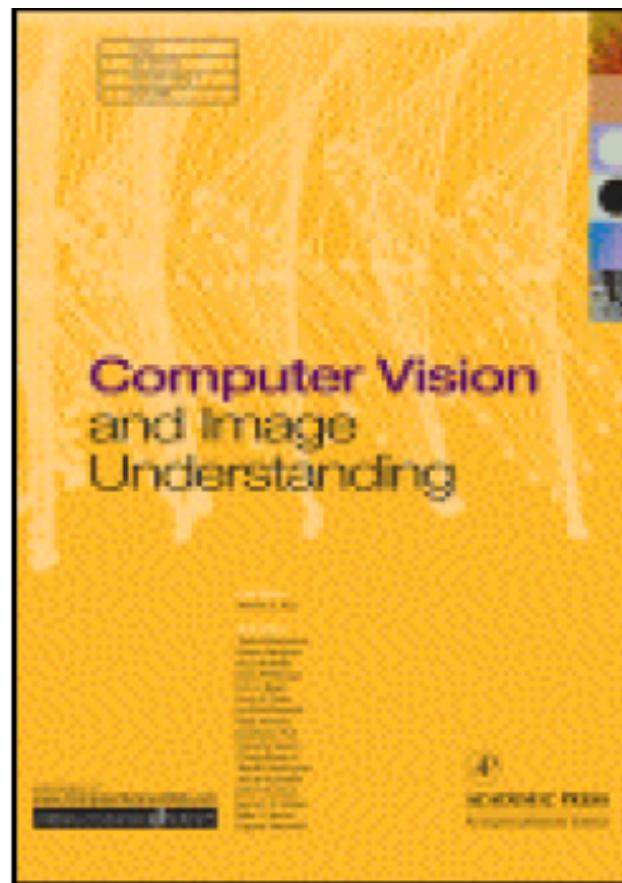




# Introduction

---

- Computer Vision and Image Understanding





# Introduction

---

- ***International Conference on Computer Vision (ICCV).***
- ***IEEE Conference on Computer Vision and Pattern Recognition (CVPR)***
- ***European Conference on Computer Vision (ECCV)***
- ***British Machine Vision Conference (BMVC)***
- ***Asian Conference on Computer Vision (ACCV)***
- ***Winter Conference on Applications of Computer Vision (WACV)***
- ***International Conference on Pattern Recognition (ICPR)***
- ***International Conference on Image Processing (ICIP)***



# Introduction

---

- [1] D. Marr, **Vision**----A computational investigation into the human representation and processing of visual information, W.H. Freeman and Company San Francisco, 1982.  
[视觉计算理论---姚国正等译, 1988.]
- [2] S.E. Palmer, **Vision Science** – Photons to Phenomenology, the MIT Press, 1999.
- [3] Forsyth, and Ponce, **Computer Vision – A Modern Approach**, (英文影印版) ,  
清华大学出版社, 2004, ISBN:7-302-07795-9。
- [4] R. Hartley, A. Zisserman, **Multiple View Geometry in Computer Vision**, 2nd Ed., Cambridge Univ. Press, 2004
- [5] Richard Szeliski, Computer Vision: **Algorithms and Applications**, Springer, 2011.
- [6] Richard O.Duda, Peter E. Hart, and David G. Stork, **Pattern Classification**, 2nd edition (英文影印版) , 北京机械工业出版社, 2004, ISBN:7-111-13687-X。
- [7] C.M. Bishop, **Pattern Recognition and Machine Learning**, Springer, 2006.
- [8]. Ian Goodfellow, Joshua Bengio, and Aaron Courville, **Deep Learning**, 2016.