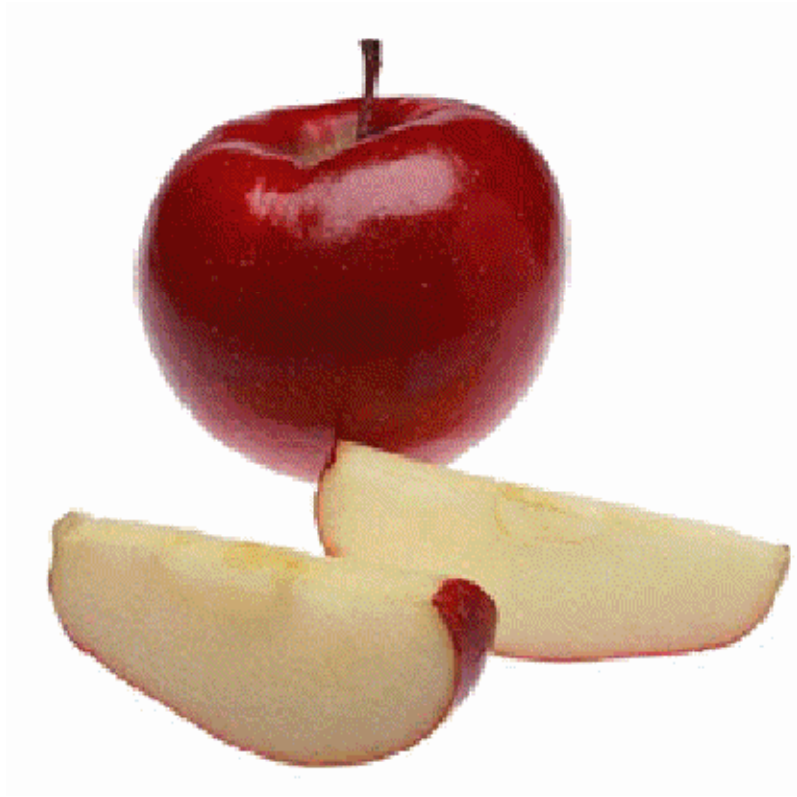


CS294-43: Visual Object and Activity Recognition

Prof. Trevor Darrell
Spring 2009

Course Content

- Contemporary vision techniques for object and activity recognition
 - instance retrieval
 - category recognition
- Comprehensive view of current best-performing methods on challenge datasets
- Readings from literature; no textbook
- Motivating applications
 - robotics
 - mobile content-based retrieval ('situated search')

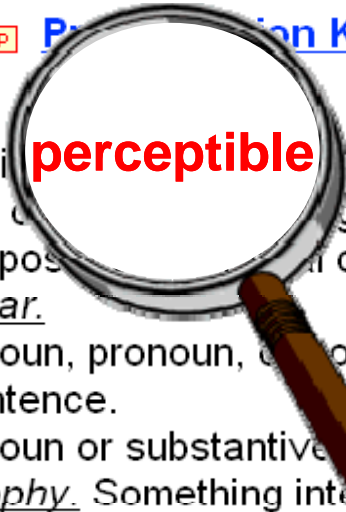


Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

ob·ject   [Pronunciation Key](#) (ˈɒbjɪkt, -jɛkt')

n.

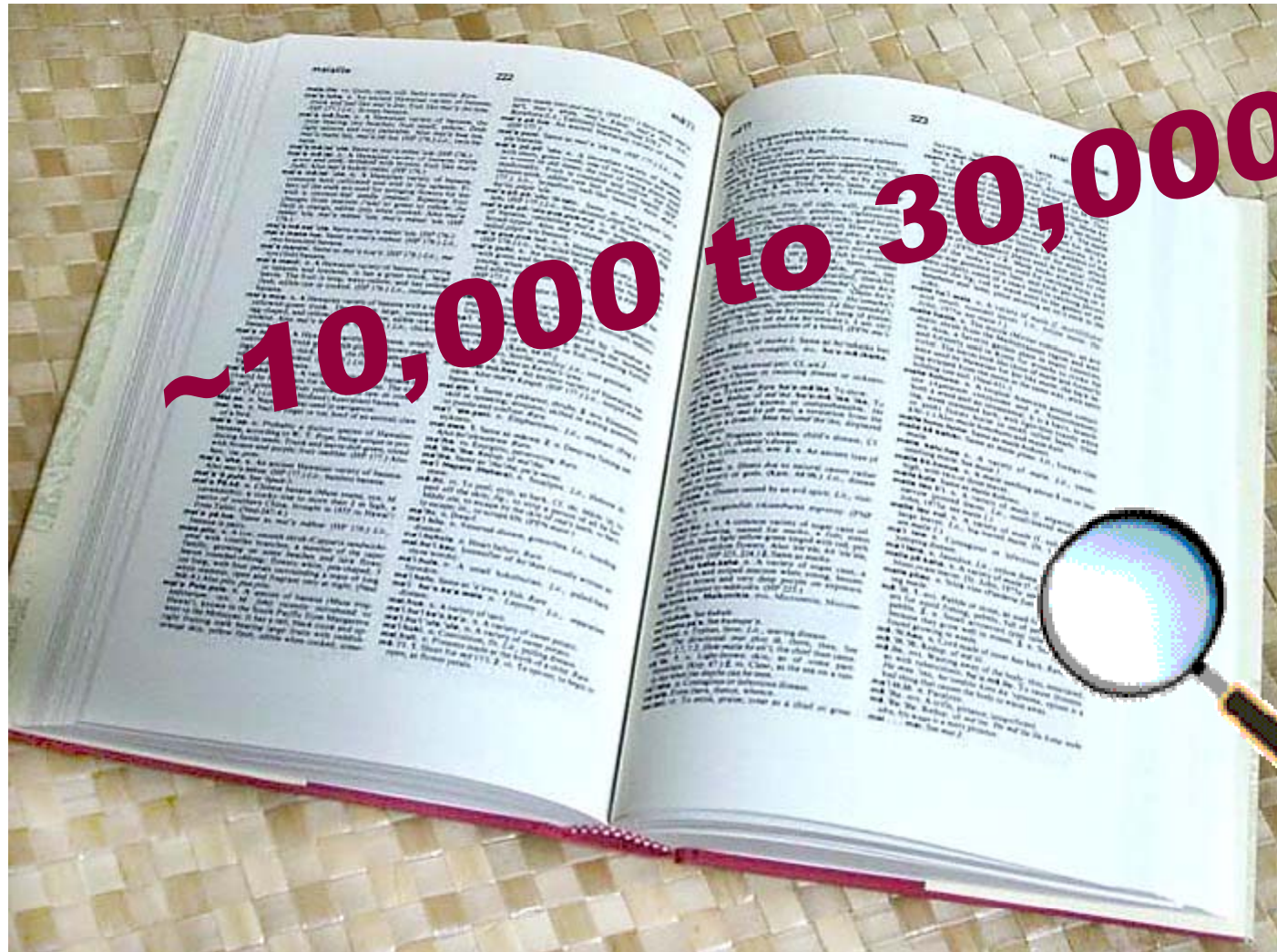
1. Something perceptible by one or more of the senses, especially sight or touch; a focus of attention: *an object of desire*.
2. A focus of activity, thought, or action: *an object of cooperation*.
3. The purpose or goal of a specific action or effort: *the object of the game*.
4. Grammar.
 - a. A noun, pronoun, or noun phrase that receives or is affected by the action of a verb within a sentence.
 - b. A noun or substantive governed by a preposition.
5. Philosophy. Something intelligible or perceptible by the mind.
6. Computer Science. A discrete item that can be selected and maneuvered, such as an onscreen graphic. In object-oriented programming, objects include data and the procedures necessary to operate on that data.





Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course **Bruegel, 1564**

How many object categories are there?



So what does object recognition involve?



Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Verification: is that a lamp?



Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Detection: are there people?



Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Identification: is that Potala Palace?



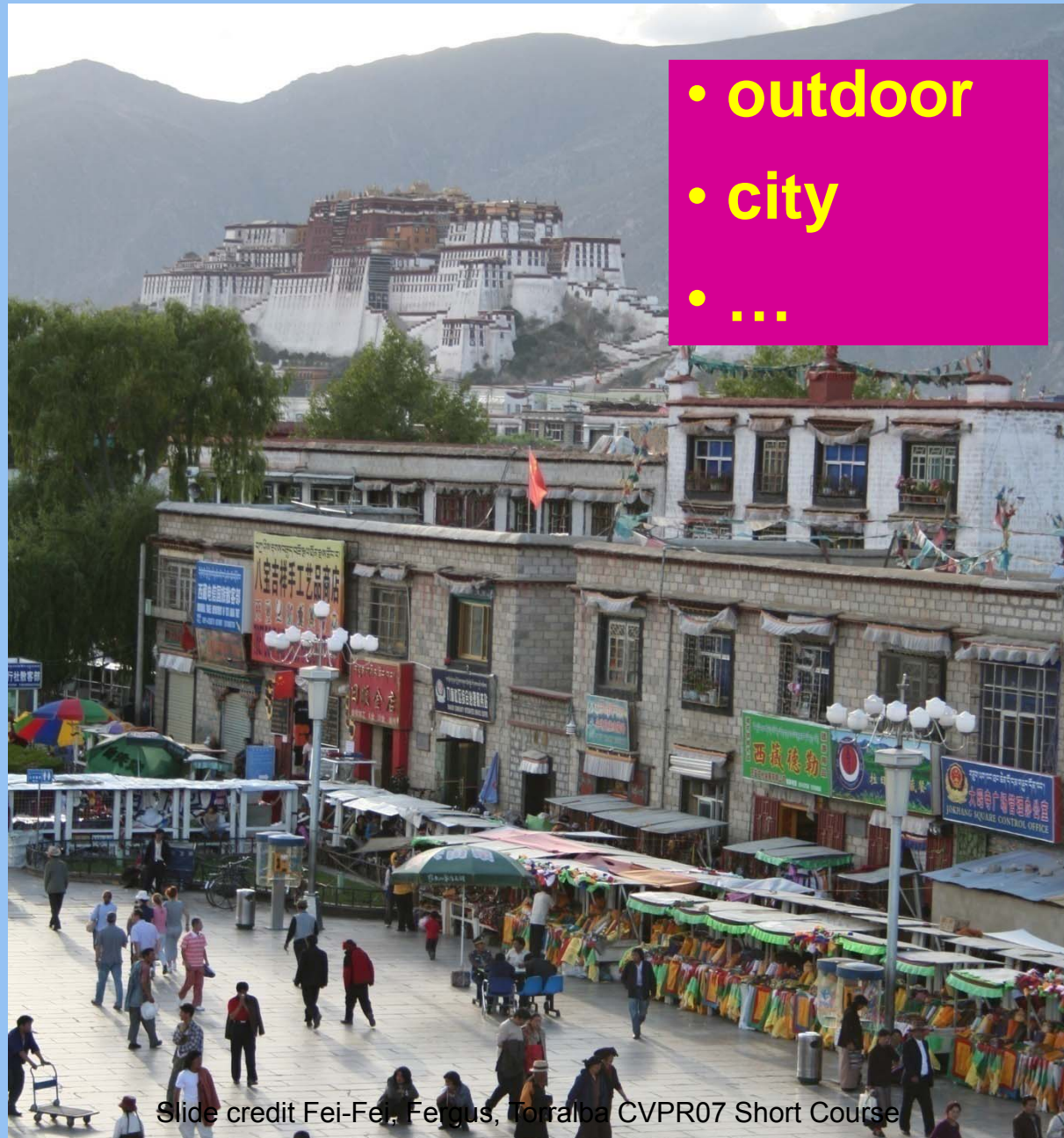
Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Object categorization



Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Scene and context categorization



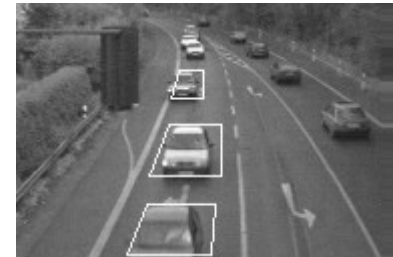
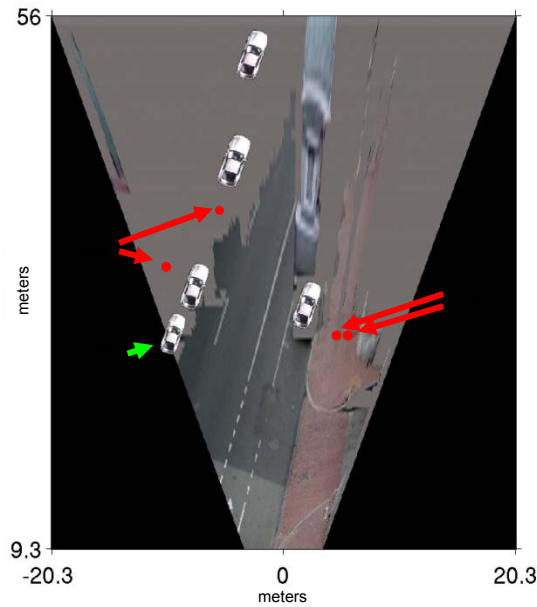
Computational photography



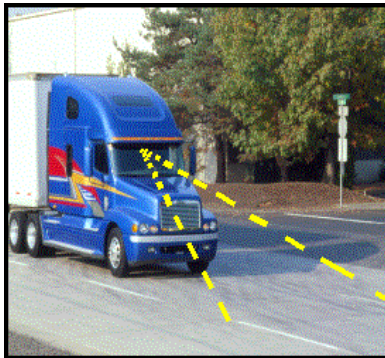
Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Assisted driving

Pedestrian and car detection



Lane detection



- Collision warning systems with adaptive cruise control,
- Lane departure warning systems,
- Rear object detection systems,

Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Improving online search

flickr ^{GAMMA}

webshots ^{beta}

Ask TM Images

Cydral TM
Image & Site Search

picsearch TM

Google TM
Image Search

altavista TM











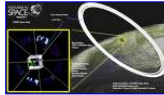

Query:
STREET

Google [web](#) [images](#) [video](#) [news](#) [maps](#) [more »](#)

street [Advanced Image Search Preferences](#)

Moderate SafeSearch is on

Images Showing: All image sizes Results 19 - 36 of about 44,200,000 for street [definition]. (0.04 seconds)

 <p>Street sweeper 345 x 352 - 17k - jpg www.town.telluride.co.us</p>	 <p>Street Maintenance 407 x 402 - 18k - jpg www.town.telluride.co.us</p>	 <p>Main Street Station 360 x 392 - 30k - jpg www.rmaonline.org</p>	 <p>SHPO Wayne Donaldson at Main Street ... 410 x 314 - 41k - jpg ohp.parks.ca.gov</p>	 <p>Lombard Street, worlds crookedest See ... 500 x 387 - 59k - jpg www.inetours.com</p>	 <p>Street Bike (BS70-4A) Details 360 x 360 - 38k - jpg bashan.en.alibaba.com</p>
 <p>Street Lamps 360 x 360 - 18k - jpg syi.en.alibaba.com</p>	 <p>Washington D.C. Laminated Street Map 500 x 500 - 114k - jpg www.dcgiftshop.com</p>	 <p>street-riders-ss-3.jpg 550 x 309 - 53k - jpg www.pspworld.com</p>	 <p>Visually Street Riders is not nearly ... 550 x 309 - 52k - jpg www.pspworld.com</p>	 <p>STREET space ring Postcards To Space ... 1000 x 563 - 87k - jpg www.postcardstospace.com</p>	 <p>17 Fleet Street 492 x 681 - 74k - jpg www.pepysdiary.com</p>

[\[More from img.alibaba.com \]](#)

Organizing photo collections

w Timeline Gift CD

Exit Search Displaying 172 pictures in 21 albums (0.003 seconds).

Starred Movies Web Albums Date Range: All Newest

Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Challenges 1: view point variation



Challenges 2: illumination



slide credit: S. Ullman

Challenges 3: occlusion



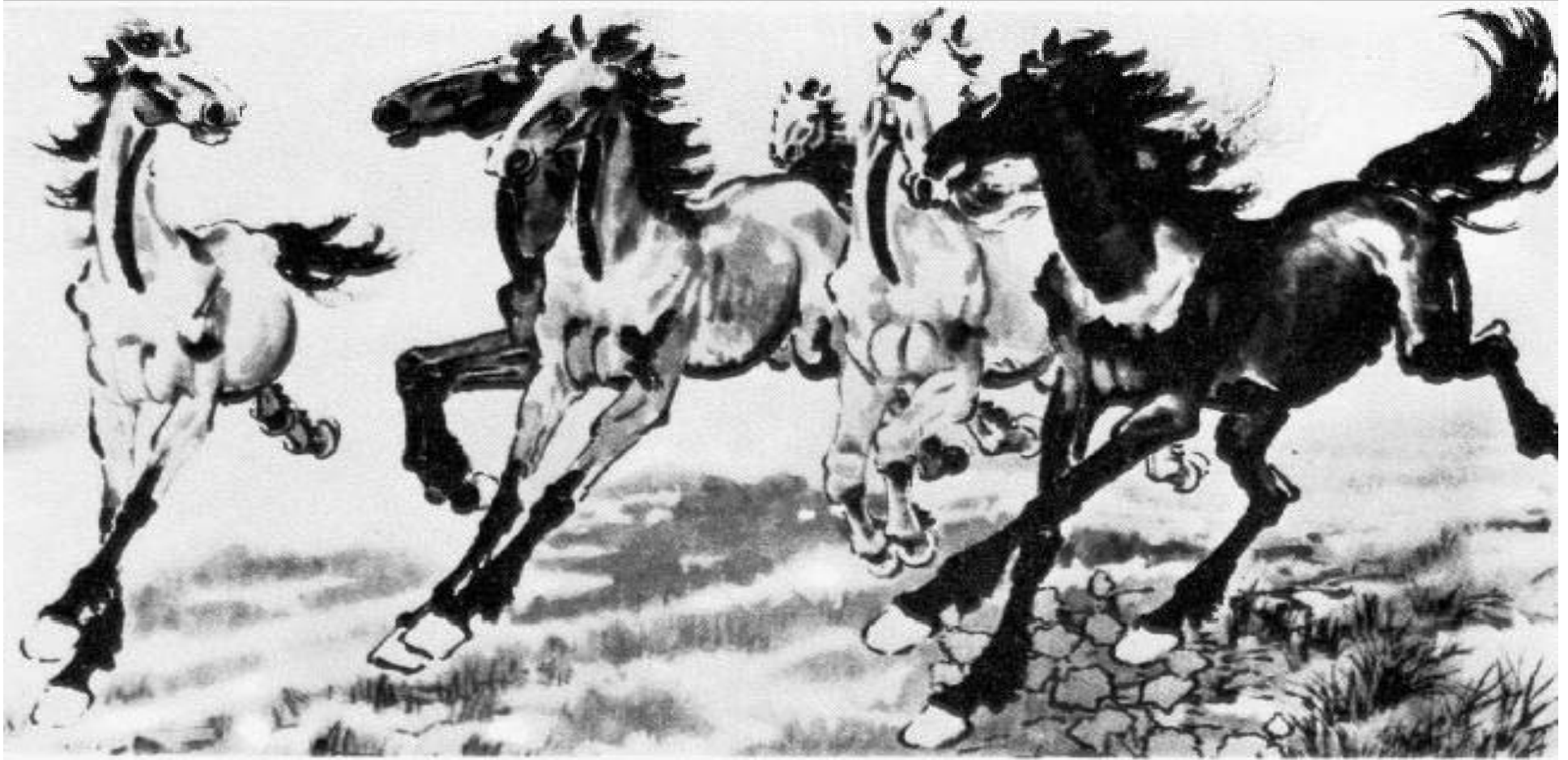
Magritte, 1957 Side Wall, Fergus, Torralba CVPR07 Short Course

Challenges 4: scale



Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Challenges 5: deformation



Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Xu, Beihong 1943

Challenges 6: background clutter



Klimt, 1913 Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

History: single object recognition



Slide credit Fei-Fei, Fergus, et al. CVPR07 Short Course

History: single object recognition



- Lowe, et al. 1999, 2003
- Mahamud and Herbert, 2000
- Ferrari, Tuytelaars, and Van Gool, 2004
- Rothganger, Lazebnik, and Ponce, 2004
- Moreels and Perona, 2005
- ...

Challenges 7: intra-class variation



Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

History: early object categorization



1 7 9 6
7 8 6 3
2 1 7 9 7 1 2
4 8 1 9 0 1 8
7 6 1 8 6 4 1
7 5 9 2 6 5 8 1 9 7
2 2 2 2 2 3 4 4 8 0
0 2 3 8 0 7 3 8 5 7
0 1 4 6 4 6 0 2 4 3
7 1 2 8 7 6 9 8 6 1



- Turk and Pentland, 1991
- Belhumeur, Hespanha, & Kriegman, 1997
- Schneiderman & Kanade 2004
- Viola and Jones, 2000



- Amit and Geman, 1999
- LeCun et al. 1998
- Belongie and Malik, 2002



- Schneiderman & Kanade, 2004
- Argawal and Roth, 2002
- Poggio et al. 1993

Slide credit Fel-Fei



~10,000 to 30,000

Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

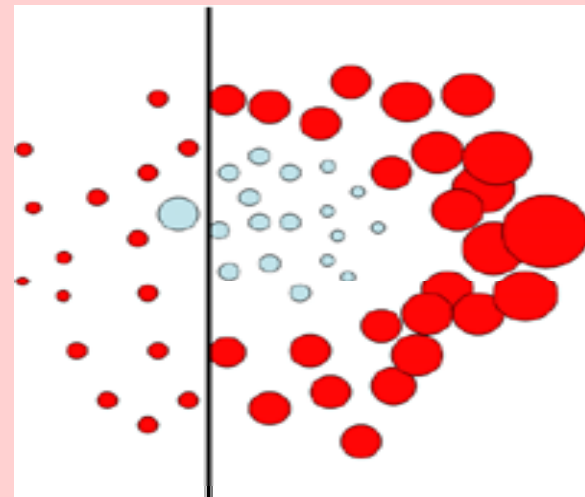
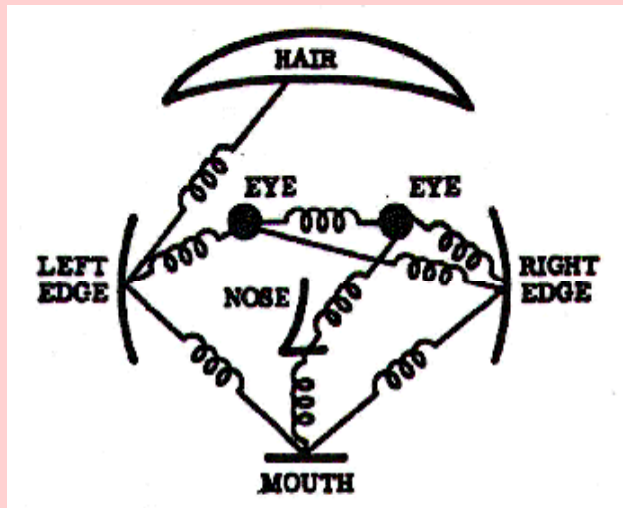


Three main issues

- Representation
 - How to represent an object category
- Learning
 - How to form the classifier, given training data
- Recognition
 - How the classifier is to be used on novel data

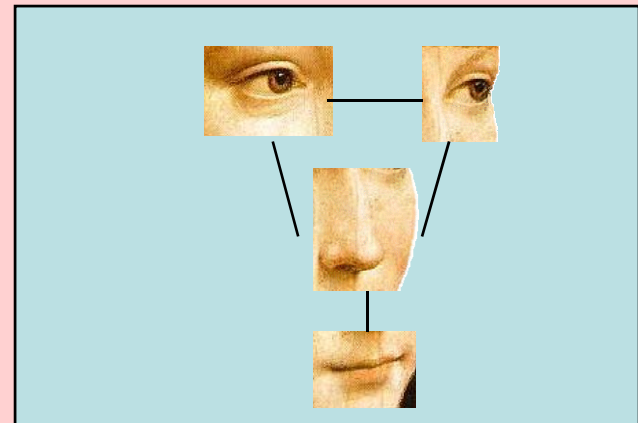
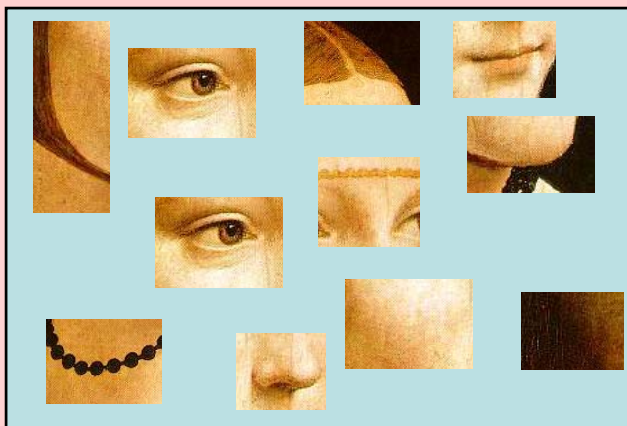
Representation

- Generative / discriminative / hybrid



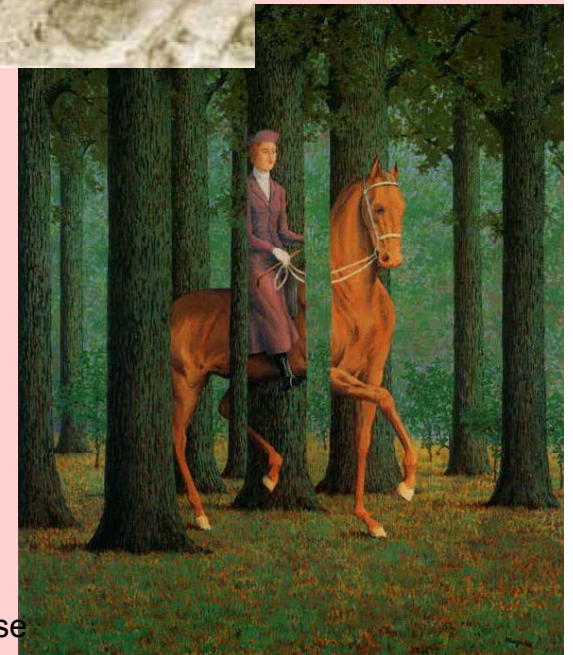
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance



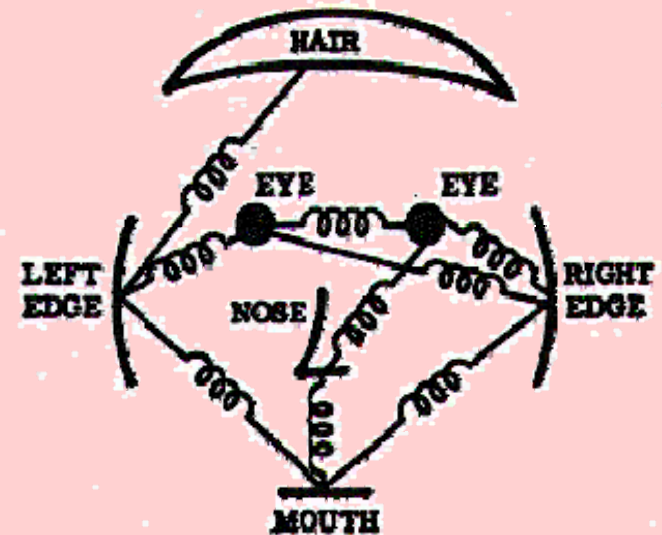
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- Invariances
 - View point
 - Illumination
 - Occlusion
 - Scale
 - Deformation
 - Clutter
 - etc.



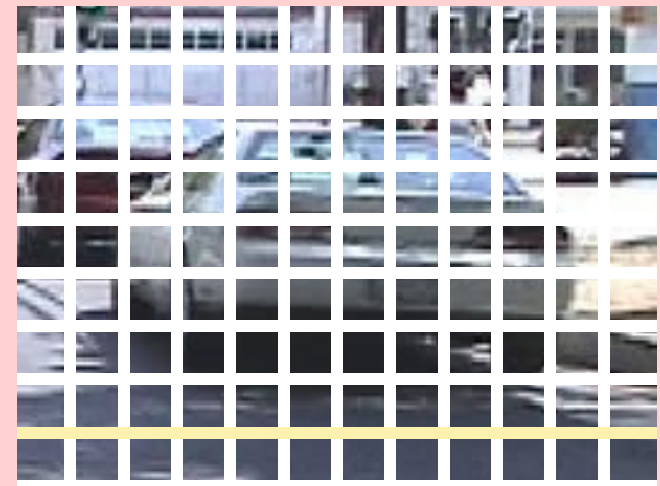
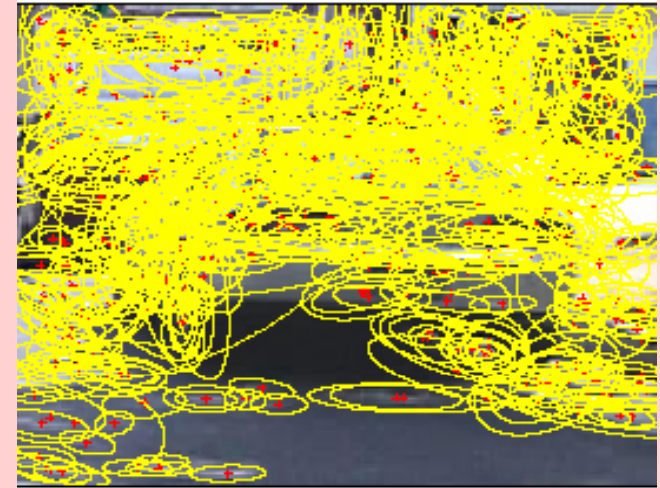
Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- invariances
- Part-based or global w/sub-window



Representation

- Generative / discriminative / hybrid
- Appearance only or location and appearance
- invariances
- Parts or global w/sub-window
- Use set of features or each pixel in image



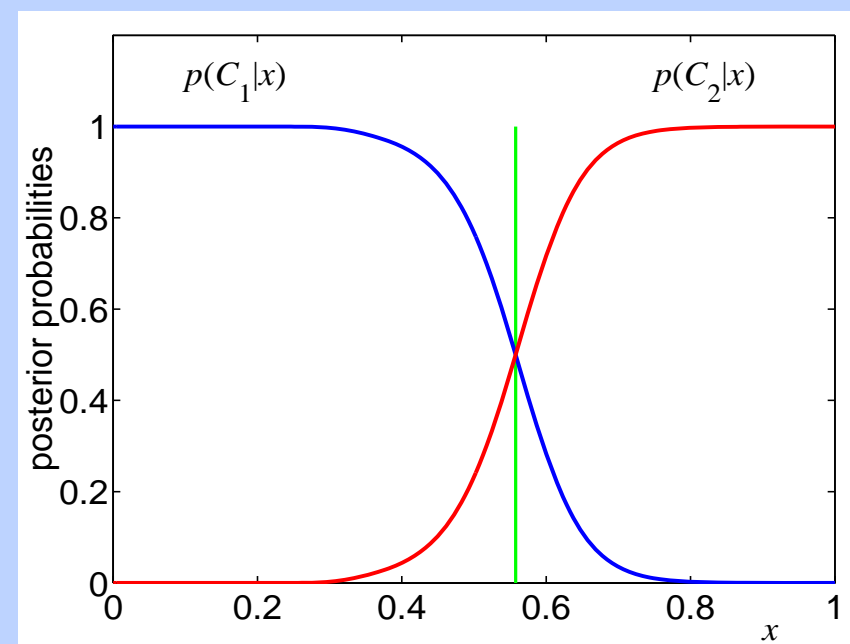
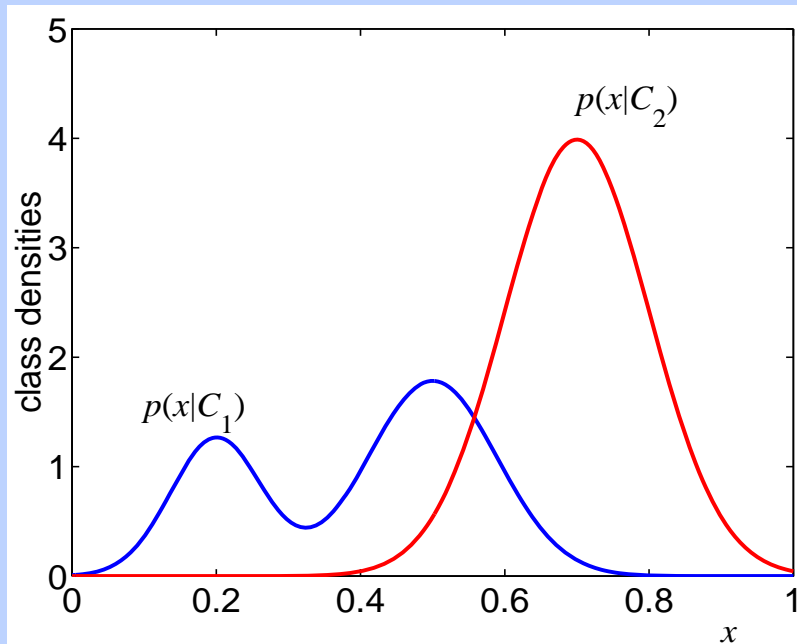
Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning



Learning

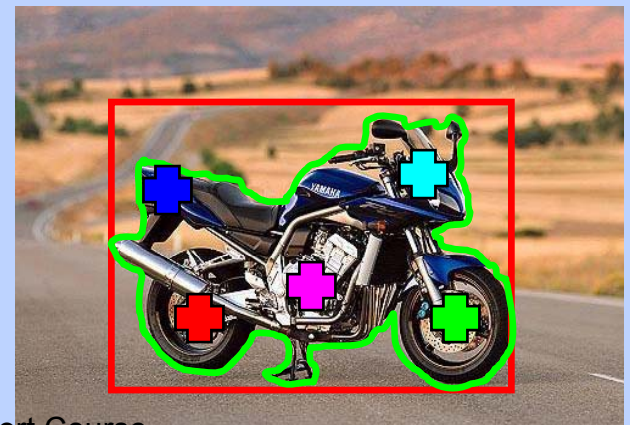
- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- Methods of training: generative vs. discriminative



Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
 - Manual segmentation; bounding box; image labels; noisy labels

Contains a motorbike



Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
 - Manual segmentation; bounding box; image labels; noisy labels
- Batch/incremental (on category and image level; user-feedback)

Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
 - What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
 - Level of supervision
 - Manual segmentation; bounding box; image labels; noisy labels
 - Batch/incremental (on category and image level; user-feedback)
 - Training images:
 - Issue of overfitting
 - Negative images for discriminative methods
- Priors

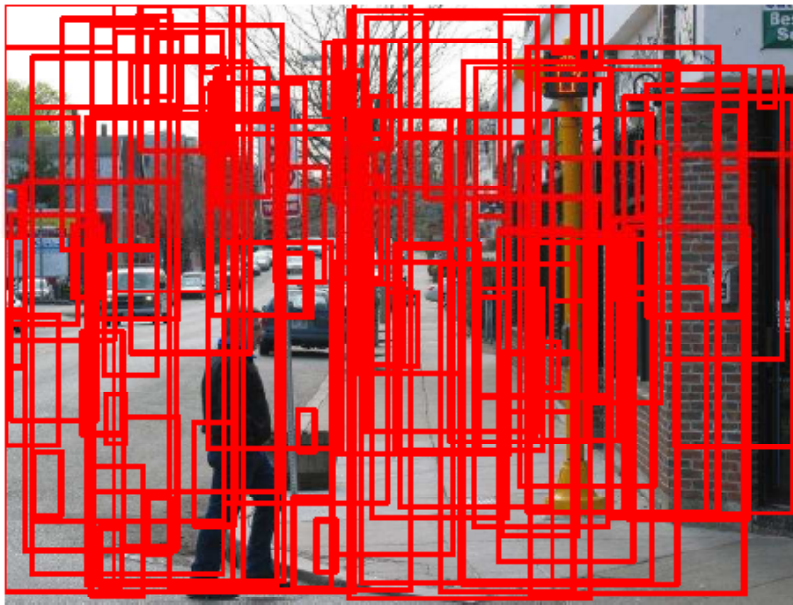
Learning

- Unclear how to model categories, so we learn what distinguishes them rather than manually specify the difference -- hence current interest in machine learning)
- What are you maximizing? Likelihood (Gen.) or performances on train/validation set (Disc.)
- Level of supervision
 - Manual segmentation; bounding box; image labels; noisy labels
- Batch/incremental (on category and image level; user-feedback)
- Training images:
 - Issue of overfitting
 - Negative images for discriminative methods
- Priors

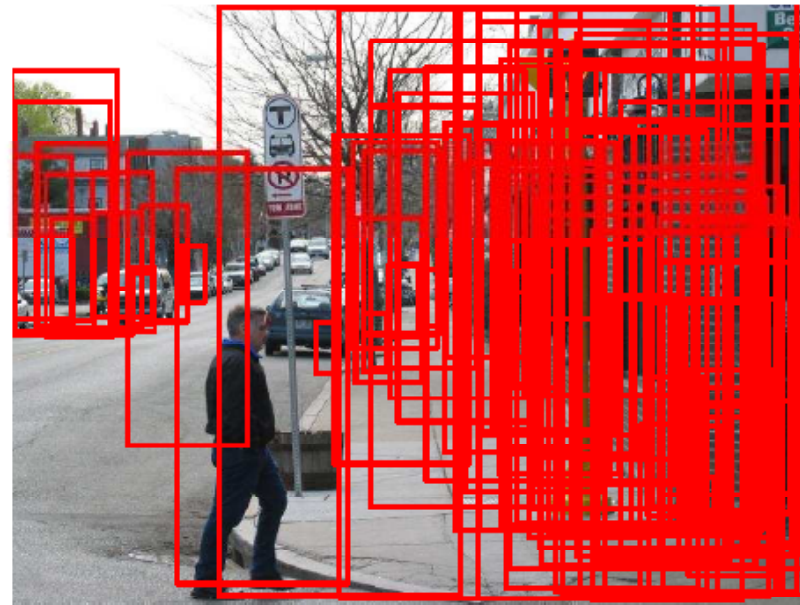
Recognition

- Scale / orientation range to search over
- Speed
- Context





(b) $P(\text{person}) = \text{uniform}$



(d) $P(\text{person} \mid \text{geometry})$



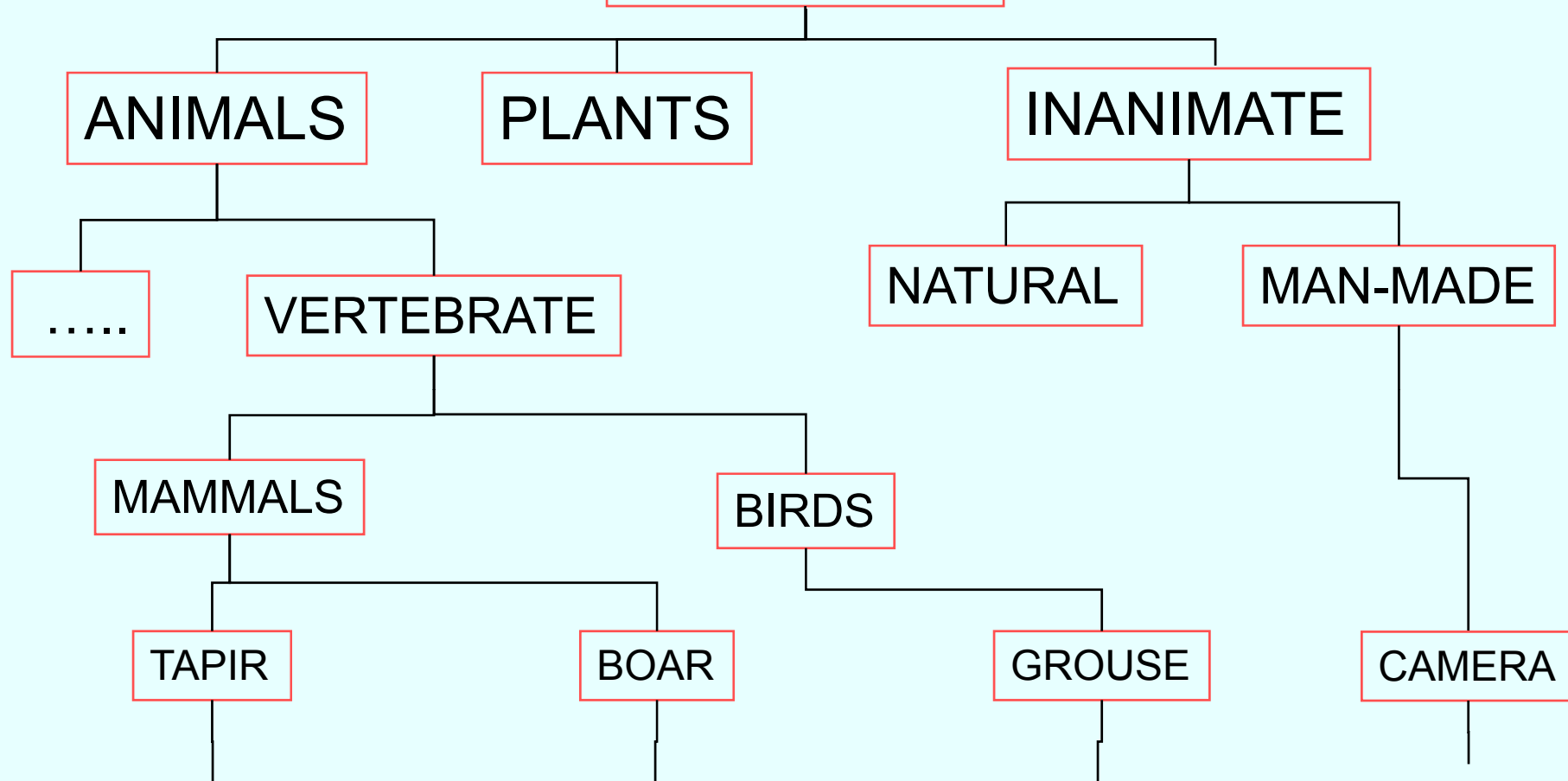
(f) $P(\text{person} \mid \text{viewpoint})$



(g) $P(\text{person} \mid \text{viewpoint, geometry})$

Slide credit: Fei-Fei Fergus, Cornell University, 2007 Short Course

OBJECTS



Slide credit Fei-Fei, Fergus, Torralba CVPR07 Short Course

Administrivia

Course Prerequisites

- Prerequisites:
 - prior Computer Vision and Machine Learning courses, or permission of instructor.
 - Advanced undergraduates allowed with permission of instructor.
- Students should already be familiar with or be willing to learn on their own:
 - basic image processing in MATLAB; Optic Flow; Edge Detection; Support Vector Machines; Gaussian Mixture Models; Hidden Markov Models, etc.

Course Requirements and Grading

- Variable units (2 or 4)
- 2 units:
 - Weekly readings (66%): in-class discussion and emailed <1 page summary of all readings **before start of class**.
 - In class presentation(s) of demo corresponding to assigned paper (34%)
- 4 units:
 - Weekly readings (33%): in-class discussion and emailed <1 page summary of all readings **before start of class**.
 - In class presentation(s) of demo corresponding to assigned paper (17%)
 - Final project (50%); proposal due March 17th, presentation and report May 5th

*Very heavy reading load:
4-7 papers per week*

Course Contacts

- Prof. Trevor Darrell
 - Soda hall office: 413
 - ICSI office: 1947 Center Street, 5th floor
 - trevor@eecs.berkeley.edu
- This course will meet once a week, Tuesday 5-7pm, in 405 Soda, *except for Feb 10th*.
- <http://groups.google.com/group/ucb-object-recognition-course>
- bSpace site: "COMPSCI 294 LEC 043 Sp09 Visual Object & Act. Rec."

Syllabus

Jan 27th – Instance recognition and retrieval

- D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, November 2004. Available: <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision. Washington, DC, USA: IEEE Computer Society, 2003. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1238663
- O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," in IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/ICCV.2007.4408891>
- N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3d," ACM Transactions on Graphics (TOG), (SIGGRAPH) 2006. <http://phototour.cs.washington.edu/>

Feb 3rd – Global features (HoG, Gist, Motion History, etc.)

- B. Schiele and J. L. Crowley, "Object recognition using multidimensional receptive field histograms," in ECCV '96: Proceedings of the 4th European Conference on Computer Vision-Volume I. London, UK: Springer-Verlag, 1996, pp. 610-619. Available: <http://citeseer.ist.psu.edu/schiele96object.html>
- A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," International Journal of Computer Vision, vol. 42, no. 3, pp. 145-175, May 2001. Available: <http://dx.doi.org/10.1023/A:1011139631724>
- A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 23, no. 3, pp. 257-267, 2001. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=910878
- A. Efros, A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," ICCV 2003, pp. 726-733 vol.2. Available: <http://dx.doi.org/10.1109/ICCV.2003.1238420>
- N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 886-893. Available: <http://dx.doi.org/10.1109/CVPR.2005.177>
- A. Yilmaz and M. Shah, "Actions sketch: A novel action representation," in CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 984-989. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1467373

Feb 17th – **Local features** (SIFT, Surf, MSER, Shape Context, Self Similarity, etc.)

- T. Lindeberg, "Feature detection with automatic scale selection," International Journal of Computer Vision, vol. 30, no. 2, pp. 79-116, November 1998. Available: <http://dx.doi.org/10.1023/A:1008045108935>
- S. Belongie, J. Malik, and J. Puzicha, "Shape context: A new descriptor for shape matching and object recognition," in NIPS, 2000, pp. 831-837. Available: <http://citeseer.ist.psu.edu/434232.html>
- J. Matas, O. Chum, U. Martin, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in Proceedings of British Machine Vision Conference, vol. 1, London, 2002, pp. 384-393. Available: <http://citeseer.ist.psu.edu/608213.html>
- K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," Int. J. Comput. Vision, vol. 60, no. 1, pp. 63-86, October 2004. Available: <http://dx.doi.org/10.1023/B:VISI.0000027790.02288.f2>
- I. Laptev, "On space-time interest points," International Journal of Computer Vision, vol. 64, no. 2-3, pp. 107-123, September 2005. Available: <http://dx.doi.org/10.1007/s11263-005-1838-7>
- E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2007.383198>

Feb 24th – **Generative approaches** (Constellation, Topic Models, etc.)

- R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2003, pp. 264-271. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1211479
- J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering object categories in image collections," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2005. <http://publications.csail.mit.edu/tmp/MIT-CSAIL-TR-2005-012.ps>
- F.-F. Li and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2. Washington, DC, USA: IEEE Computer Society, 2005, pp. 524-531. Available: <http://dx.doi.org/10.1109/CVPR.2005.16>
- J. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," International Journal of Computer Vision. 79(3): 299-318. 2008 Available: <http://dx.doi.org/10.1007/s11263-007-0122-4>
- P. Moreels and P. Perona, "A probabilistic cascade of detectors for individual object recognition," European Conference on Computer Vision , vol III, pp.426-439, 2008 Available: http://dx.doi.org/10.1007/978-3-540-88690-7_32
- E. Sudderth, A. Torralba, W. Freeman, and A. Willsky, "Describing visual scenes using transformed objects and parts," International Journal of Computer Vision, vol. 77, no. 1, pp. 291-330, May 2008. Available: <http://dx.doi.org/10.1007/s11263-007-0069-5>

March 3rd – Voting and Indexing techniques (ISM, k-NN, LSH, Random Forests, Metric Learning, etc.)

- B. Leibe, A. Leonardis, and B. Schiele, "An implicit shape model for combined object categorization and segmentation," In ECCV workshop on statistical learning in computer vision 2006, pp. 508-524. Available: http://dx.doi.org/10.1007/11957959_26
- J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587503>
- A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning globally-consistent local distance functions for shape-based image retrieval and classification," in Proceedings of IEEE 11th International Conference on Computer Vision, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/ICCV.2007.4408839>
- M. Ozuysal, P. Fua, and V. Lepetit, "Fast keypoint recognition in ten lines of code," in Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2007.383123>
- P. Jain, B. Kulis, and K. Grauman, "Fast image search for learned metrics," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587841>
- A. Torralba, R. Fergus, and Y. Weiss, "Small codes and large image databases for recognition," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587633>

March 10th – Discriminative approaches

- C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints," in ECCV International Workshop on Statistical Learning in Computer Vision, 2004. Available: http://www.xrce.xerox.com/Publications/Attachments/2004-010/2004_010.pdf
- B. Caputo, C. Wallraven, and M. E. Nilsback, "Object categorization via local kernels," in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, vol. 2, 2004, pp. 132-135 Vol.2. Available: <http://dx.doi.org/10.1109/ICPR.2004.1334079>
- H. Zhang, A. C. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2126-2136. Available: <http://dx.doi.org/10.1109/CVPR.2006.301>
- A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, "Active learning with gaussian processes for object categorization," in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/ICCV.2007.4408844>
- P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska, June 2008., June 2008. Available: <http://www.ics.uci.edu/~dramanan/papers/latent.pdf>
- Y. Wang and G. Mori, "Learning a Discriminative Hidden Part Model for Human Action Recognition", Advances in Neural Information Processing Systems (NIPS), 2008; <http://www.sfu.ca/~ywang12/papers/nips.pdf>

March 17th – Correspondence and Pyramid-based techniques (EMD, PMK, SPMK, SPK, etc.)

- A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1. Washington, DC, USA: IEEE Computer Society, 2005, pp. 26-33. Available: <http://dx.doi.org/10.1109/CVPR.2005.320>
- K. Grauman and T. Darrell, "The pyramid match kernel: discriminative classification with sets of image features," ICCV vol. 2, 2005, pp. 1458-1465 Vol. 2. Available: <http://dx.doi.org/10.1109/ICCV.2005.239>
- K. Grauman and T. Darrell, "Approximate correspondences in high dimensions," in In NIPS, vol. 2006, 2006. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.90.3400>
- S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," CVPR vol. 2, 2006, pp. 2169-2178. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1641019
- A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval. New York, NY, USA: ACM Press, 2007, pp. 401-408. Available: <http://dx.doi.org/10.1145/1282280.1282340>
- S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587630>

March 31st – Category Discovery from the Web

- R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning object categories from google's image search," vol. 2, 2005, pp. 1816-1823 Vol. 2. Available: <http://dx.doi.org/10.1109/ICCV.2005.142>
- L.-J. Li, G. Wang, and L. Fei-Fei, "Optimol: automatic online picture collection via incremental model learning," in Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007, pp. 1-8. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4270073
- F. Schroff, A. Criminisi, and A. Zisserman, "Harvesting image databases from the web," in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/ICCV.2007.4409099>
- K. Saenko and T. Darrell, "Unsupervised Learning of Visual Sense Models for Polysemous Words". Proc. NIPS, December 2008, Vancouver, Canada. http://people.csail.mit.edu/saenko/saenko_nips08.pdf

April 7th – Kernel Combination, Segmentation, and Structured Output

- M. Varma and D. Ray, "Learning the discriminative power-invariance trade-off," in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/ICCV.2007.4408875>
- Q. Yuan, A. Thangali, V. Ablavsky, and S. Sclaroff, "Multiplicative kernels: Object detection, segmentation and pose estimation," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587739>
- A. Bosch, A. Zisserman, and X. Munoz, "Image classification using rois and multiple kernel learning," IJCV 2008, 2008. Available: http://eia.udg.es/%7Eaboschr/Publicacions/bosch08a_preliminary.pdf
- C. Pantofaru, C. Schmid, and M. Hebert, "Object recognition by integrating multiple image segmentations," CVPR 2008, pp. 481-494. Available: http://dx.doi.org/10.1007/978-3-540-88690-7_36
- M. B. Blaschko and C. H. Lampert, "Learning to localize objects with structured output regression," in ECCV 2008. Lecture Notes in Computer Science, D. A. Forsyth, P. H. S. Torr, A. Zisserman, D. A. Forsyth, P. H. S. Torr, and A. Zisserman, Eds., vol. 5302. Springer, 2008, pp. 2-15. Available: http://dx.doi.org/10.1007/978-3-540-88682-2_2

April 14th – Image Context

- A. Torralba, K. P. Murphy, and W. T. Freeman, "Contextual models for object detection using boosted random fields," in Advances in Neural Information Processing Systems 17 (NIPS), 2005, pp. 1401-1408. .
<http://dspace.mit.edu/handle/1721.1/6740>
- D. Hoiem, A. A. Efros, and M. Hebert, "Putting objects in perspective," in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2, 2006, pp. 2137-2144. Available: <http://dx.doi.org/10.1109/CVPR.2006.232>
- L.-J. Li and L. Fei-Fei, "What, where and who? classifying events by scene and object recognition," in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1-8. Available:
<http://dx.doi.org/10.1109/ICCV.2007.4408872>
- S. Gould, J. Arfvidsson, A. Kaehler, B. Sapp, M. Messner, G. R. Bradski, P. Baumstarck, S. Chung, A. Y. Ng: Peripheral-Foveal Vision for Real-time Object Recognition and Tracking in Video. IJCAI 2007: 2115-2121
<http://www.stanford.edu/~sgould/papers/ijcai07-peripheralfoveal.pdf>
- G. Heitz and D. Koller, "Learning spatial context: Using stuff to find things," in ECCV 2008, pp. 30-43. Available: http://dx.doi.org/10.1007/978-3-540-88682-2_4
- Y. Li and R. Nevatia, "Key object driven multi-category object recognition, localization and tracking using spatio-temporal context," in ECCV 2008, pp. 409-422. Available:
http://dx.doi.org/10.1007/978-3-540-88693-8_30

April 21st – Shared Structures (Features, Parts)

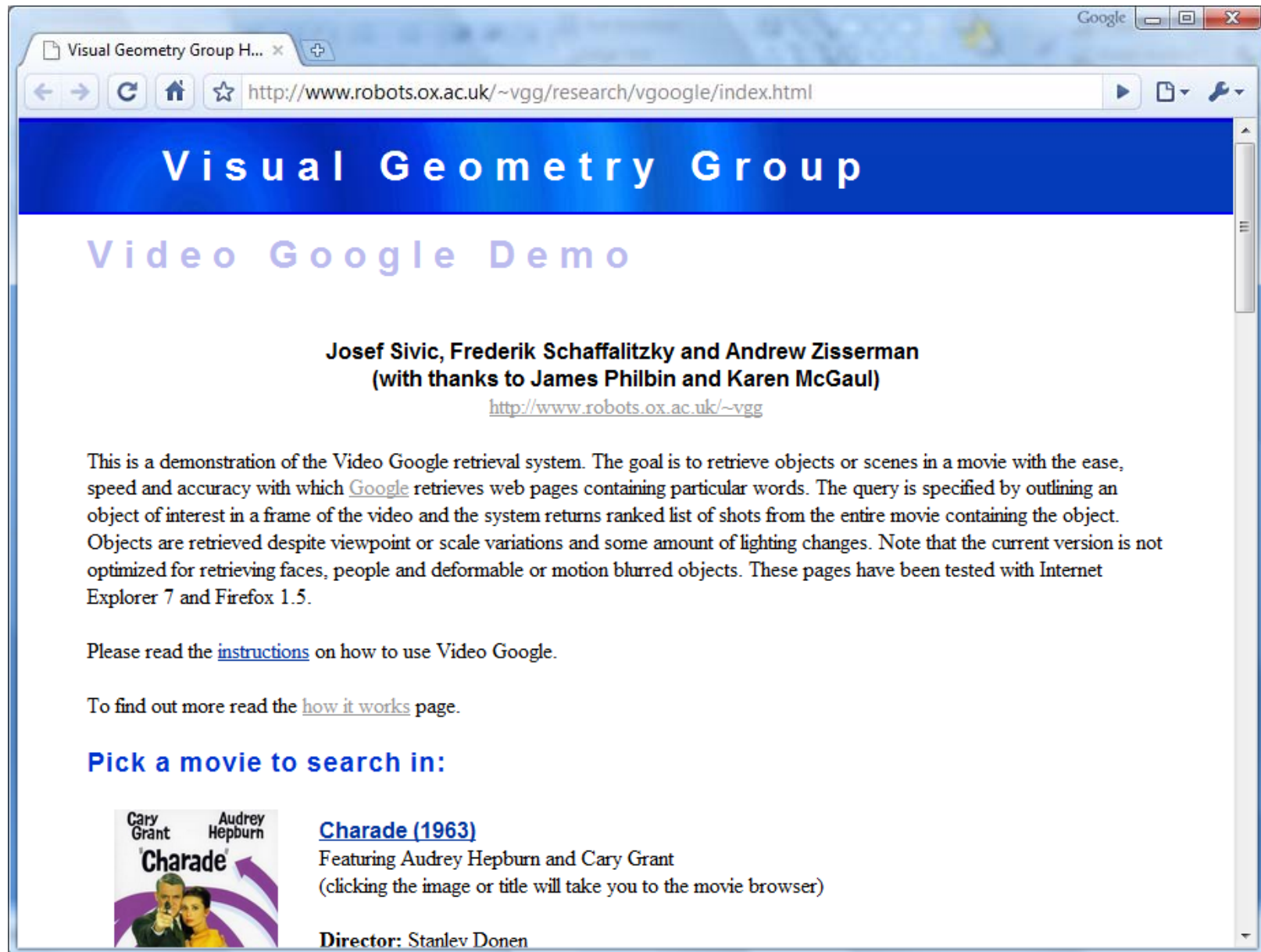
- A. Quattoni, M. Collins, and T. Darrell, "Transfer learning for image classification with sparse prototype representations," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587637>
- A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing visual features for multiclass and multiview object detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 5, pp. 854-869, 2007. Available: <http://dx.doi.org/10.1109/TPAMI.2007.1055>
- S. Fidler and A. Leonardis, "Towards scalable representations of object categories: Learning a hierarchy of parts," in Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2007.383269>
- T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Object recognition with cortex-like mechanisms. PAMI, 29(3):411–426, 2007. <http://cbcl.mit.edu/publications/ps/serre-wolf-poggio-PAMI-07.pdf>

April 28th – Hierarchy and Taxonomy Discovery

- A. Zweig and D. Weinshall, "Exploiting object hierarchy: Combining models from different category levels," in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/ICCV.2007.4409064>
- G. Griffin and P. Perona, "Learning and using taxonomies for fast visual categorization," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587410>
- J. Sivic, B. C. Russell, A. Zisserman, W. T. Freeman, and A. A. Efros, "Unsupervised discovery of visual object class hierarchies," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587622>
- M. Marszałek and C. Schmid, "Constructing category hierarchies for visual recognition," in ECCV 2008, pp. 479-491. Available: http://dx.doi.org/10.1007/978-3-540-88693-8_35
- E. Bart, I. Porteous, P. Perona, and M. Welling, "Unsupervised learning of visual taxonomies," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8. Available: <http://dx.doi.org/10.1109/CVPR.2008.4587620>

Demo previews for next
week...

Video Google



Visual Geometry Group H... x

http://www.robots.ox.ac.uk/~vgg/research/vgoogle/index.html

Visual Geometry Group

Video Google Demo


Josef Sivic, Frederik Schaffalitzky and Andrew Zisserman
(with thanks to James Philbin and Karen McGaul)
<http://www.robots.ox.ac.uk/~vgg>

This is a demonstration of the Video Google retrieval system. The goal is to retrieve objects or scenes in a movie with the ease, speed and accuracy with which [Google](#) retrieves web pages containing particular words. The query is specified by outlining an object of interest in a frame of the video and the system returns ranked list of shots from the entire movie containing the object. Objects are retrieved despite viewpoint or scale variations and some amount of lighting changes. Note that the current version is not optimized for retrieving faces, people and deformable or motion blurred objects. These pages have been tested with Internet Explorer 7 and Firefox 1.5.

Please read the [instructions](#) on how to use Video Google.

To find out more read the [how it works](#) page.

Pick a movie to search in:

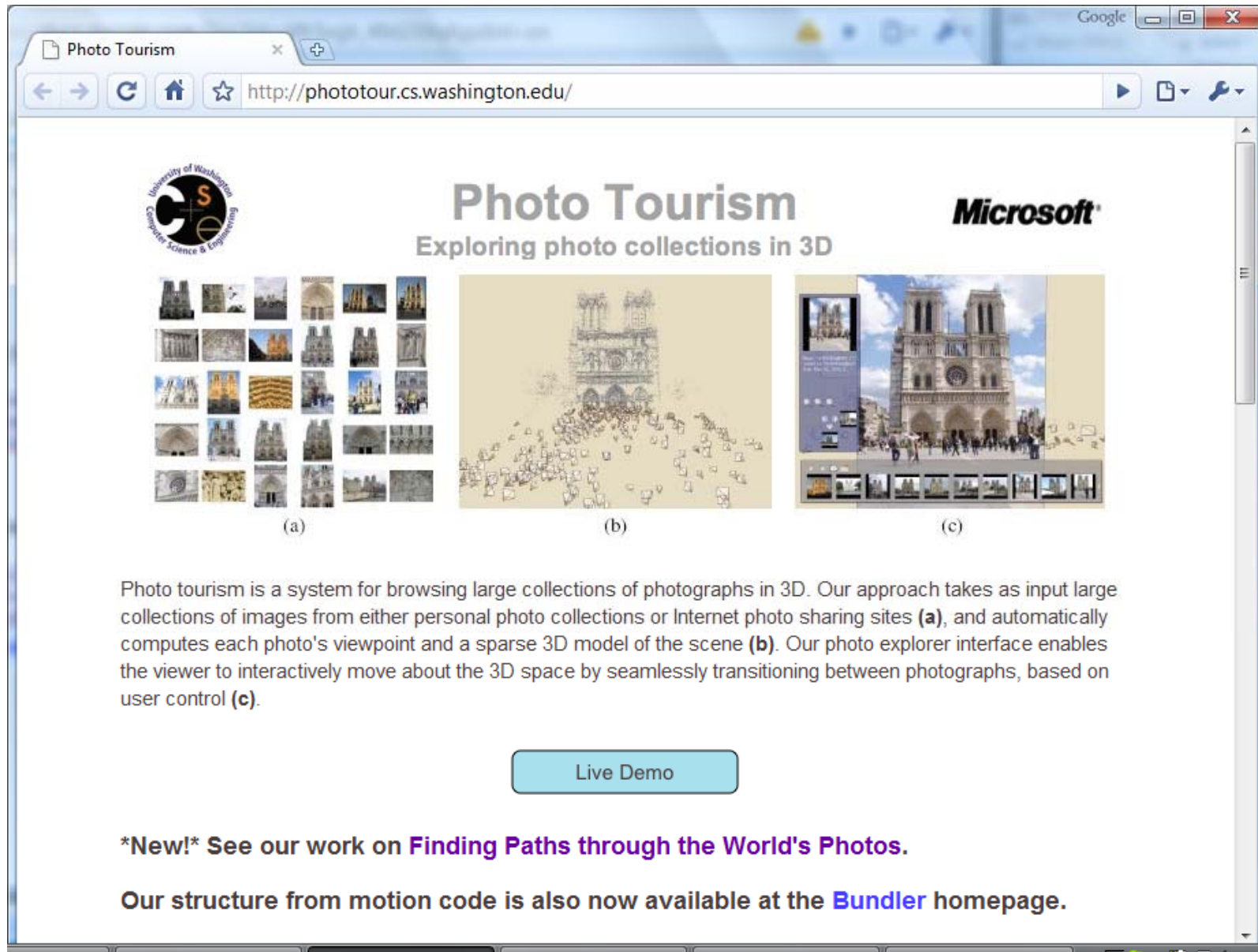


Cary Grant Audrey Hepburn
Charade

[Charade \(1963\)](#)
Featuring Audrey Hepburn and Cary Grant
(clicking the image or title will take you to the movie browser)

Director: Stanley Donen

Photo Tourism



The screenshot shows a web browser window with the address bar displaying <http://phototour.cs.washington.edu/>. The page features the University of Washington logo on the left, the title "Photo Tourism" in the center, and the Microsoft logo on the right. Below the title is the subtitle "Exploring photo collections in 3D". Three main visual elements are shown: (a) a grid of 24 small photographs of Notre-Dame de Paris; (b) a 3D wireframe model of the cathedral's facade with a path of small photo icons leading to it; and (c) a 3D interactive interface showing a large photo of the cathedral with a navigation bar at the bottom. Below these images is a paragraph of text explaining the system, followed by a "Live Demo" button and two promotional lines.

Photo Tourism

Exploring photo collections in 3D

Microsoft

(a) (b) (c)

Photo tourism is a system for browsing large collections of photographs in 3D. Our approach takes as input large collections of images from either personal photo collections or Internet photo sharing sites **(a)**, and automatically computes each photo's viewpoint and a sparse 3D model of the scene **(b)**. Our photo explorer interface enables the viewer to interactively move about the 3D space by seamlessly transitioning between photographs, based on user control **(c)**.

[Live Demo](#)

New! See our work on [Finding Paths through the World's Photos](#).

Our structure from motion code is also now available at the [Bundler](#) homepage.

Nokia Point and Tell...

<http://conversations.nokia.com/home/2008/09/point-and-fin-1.html>

The screenshot shows a web browser window displaying the Nokia Conversations website. The browser's address bar shows the URL <http://conversations.nokia.com/home/2008/09/point-and-fin-1.html>. The website header features the "Nokia Conversations" logo with the tagline "Stories from around the neighborhood". A navigation menu includes "Products & Services", "Design", "Future Technologies", "Environment", "Ideas & Opinions", and "Our Business".

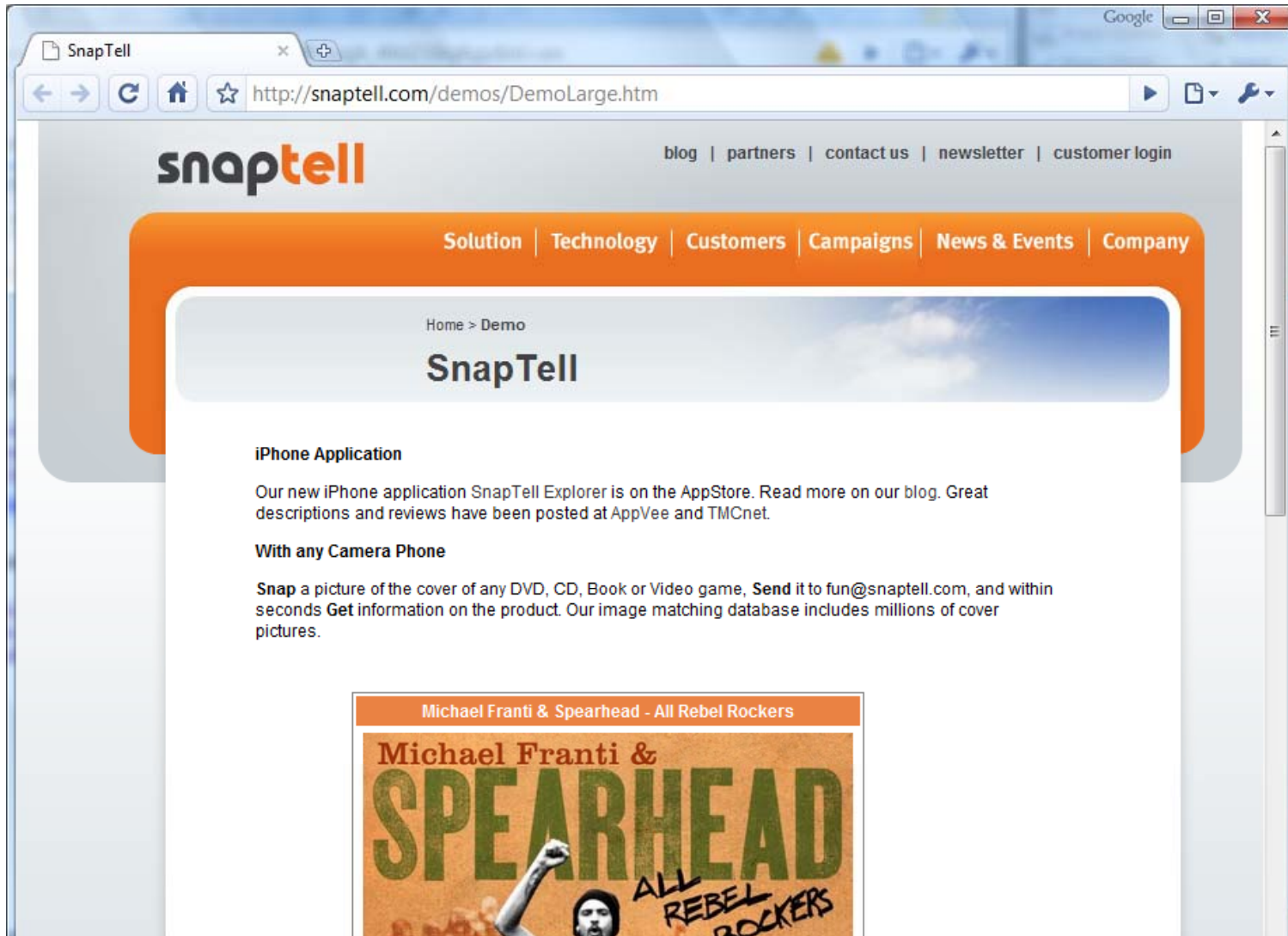
On the left side, there is a search bar labeled "Search this site" and a "Go!" button. Below it is an "Editor's desk" section with a list of links: "About us", "The ways to follow us", "Suggest a topic", "Site messages", and "Nokia N96 World Tour - Center Stage". A "Hot topics" section follows with links like "Your music, where you want it", "Days of Change", "Recycling kicks off in India", "Nokia 5800 XpressMusic faces some tough tests", and "Getting what you pay for in free online services". At the bottom left is a "Recent Comments" section with a link to "Farhan on Indoor".

The main content area is dated "24 September 2008" and features a video titled "Point and find uncovered (video)". The text describes the arrival of Nokia's Point and Find services in Espoo, Finland, and includes a link to a video showing the real-world application. A thumbnail for the video shows the text "12.12.08 IS THE DAY THE EARTH STOOD STILL". Below the text is a video player showing a person sitting on a sofa.

On the right side, there is a "Video of the day" section featuring a "Mikko Rieger interview" with a video player. Below that is an "Our picks" section with a list of links: "Nokia lanza el N79 Eco (prescinde del cargador) - Engadget en Espanol (in Spanish)", "Nokia N79 - Miljövänligare förpackad N79 säljs i Storbritannien", "Nokia Online Shop UK - Phones | All Phones | Nokia N79", "Nokia N79 Eco - buy a phone without a charger", "Nokia Europe - Caring for your mobile phone", and "Google drops Jaiku. can".

SnapTell

<http://snaptell.com/demos/DemoLarge.htm>



The screenshot shows a web browser window displaying the SnapTell website. The browser's address bar shows the URL <http://snaptell.com/demos/DemoLarge.htm>. The website features a navigation menu with links for "blog", "partners", "contact us", "newsletter", and "customer login". Below this is a secondary menu with links for "Solution", "Technology", "Customers", "Campaigns", "News & Events", and "Company". The main content area has a breadcrumb trail "Home > Demo" and a large heading "SnapTell".


iPhone Application

Our new iPhone application SnapTell Explorer is on the AppStore. Read more on our blog. Great descriptions and reviews have been posted at AppVee and TMCnet.

With any Camera Phone

Snap a picture of the cover of any DVD, CD, Book or Video game, **Send** it to fun@snaptell.com, and within seconds **Get** information on the product. Our image matching database includes millions of cover pictures.

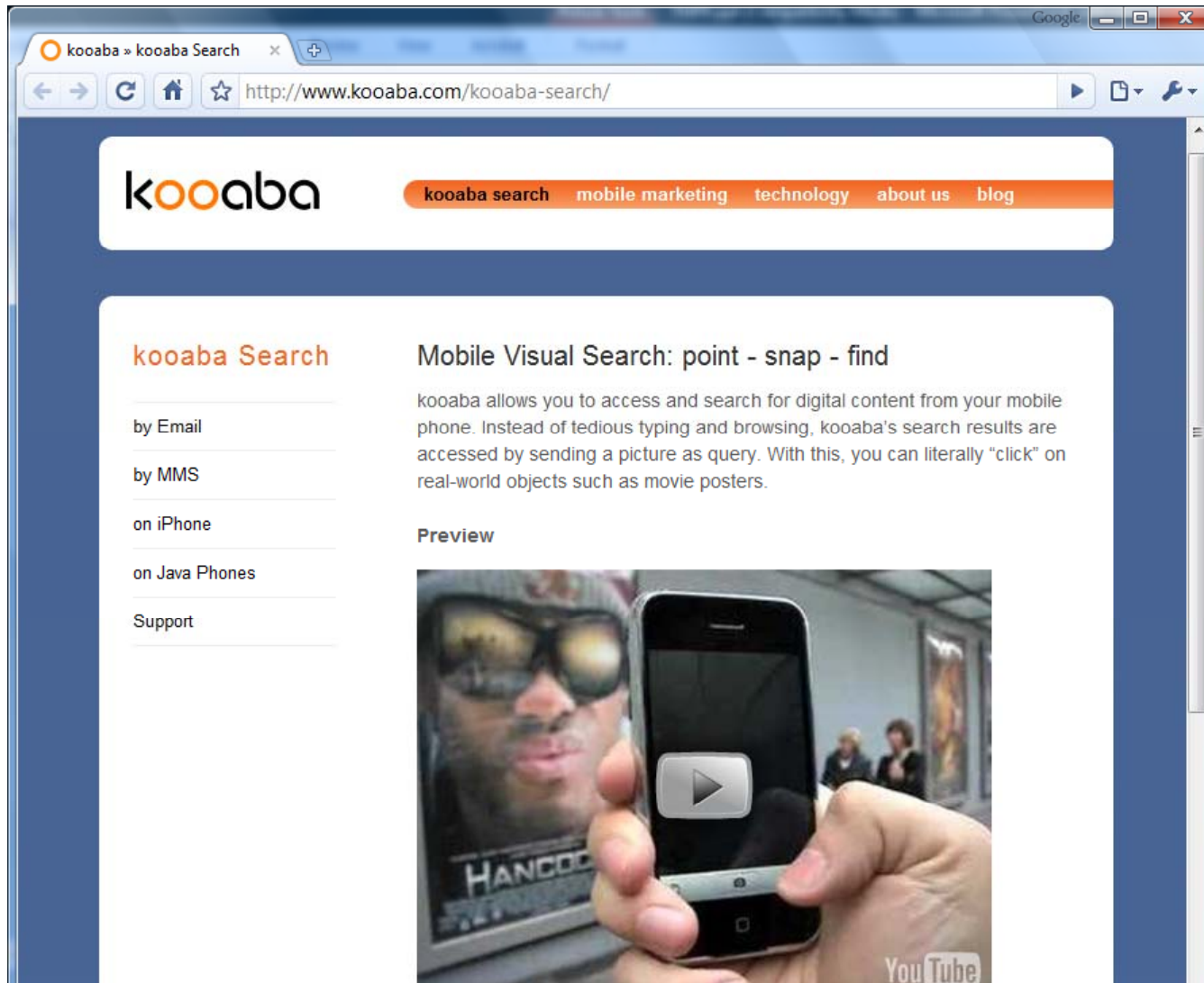
Michael Franti & Spearhead - All Rebel Rockers



The album cover for "All Rebel Rockers" by Michael Franti & Spearhead is displayed. It features a man with a raised fist and the text "ALL REBEL ROCKERS" in a stylized font.

Kooaba

<http://www.kooaba.com/kooaba-search/>



The screenshot shows a web browser window displaying the Kooaba website. The browser's address bar shows the URL <http://www.kooaba.com/kooaba-search/>. The website's header features the Kooaba logo and a navigation menu with links for "kooaba search", "mobile marketing", "technology", "about us", and "blog".

The main content area is titled "kooaba Search" and includes a sidebar with links for "by Email", "by MMS", "on iPhone", "on Java Phones", and "Support". The main text area is titled "Mobile Visual Search: point - snap - find" and contains the following text:

kooaba allows you to access and search for digital content from your mobile phone. Instead of tedious typing and browsing, kooaba's search results are accessed by sending a picture as query. With this, you can literally "click" on real-world objects such as movie posters.

Below the text is a "Preview" section featuring a video player. The video shows a hand holding an iPhone in front of a movie poster for " Hancock ". The video player has a play button in the center and a "YouTube" logo in the bottom right corner.

PhotoQA

http://poq.csail.mit.edu:3000/pdf_book/query

The screenshot shows a web browser window with the address bar containing the URL http://poq.csail.mit.edu:3000/pdf_book/query_result?query_id=890&search_terms=windows. The page is divided into a 'Query' section and a 'Result' section.

Query: The 'Date and Time' dialog box is shown with the date set to Tuesday, January 20, 2008, and the time set to 8:00 AM. The 'Time zone' is set to (GMT-08:00) Pacific Time (US & Canada). Below the dialog box, the keywords 'windows' are entered in a text box.

Result: The search results are displayed in a grid. The first result is titled '77: The Complete Guide to Windows Server 2008 Addison Wes..., page 84'. The snippet shows a section titled 'USER ACCESS CONTROL' with a page number of 53. The text discusses a function in Windows Vista that requires administrator privileges. A note states that the shield icon on the 'Change Date and Time' button indicates that administrator credentials are needed. Below the text is a small image of the 'Date and Time' dialog box with a shield icon on the 'Change Date and Time' button. The caption for this image reads: 'FIGURE 7-4 The shield icon on the Change Date and Time button shows that administrator credentials are needed to perform the action.' The second result is titled '35: Windows Vista Just the Steps For Dummies Computer Tech, page 18'. The snippet shows a section titled 'Set the Date and Time' with a list of steps: 1. Press the Windows key on your keyboard to display the taskbar if it isn't visible. 2. Right-click the Date/Time display on the far right of the taskbar and then choose Adjust Date/Time from the shortcut menu that appears. 3. Click the Change Date and Time (see figure 1-6) button and in the Date and Time Settings dialog box click another date on the calendar. Enter a new time in the 'Time' box. The snippet also includes a small image of the 'Date and Time' dialog box.

Readings for next class (Jan 27th) – Instance recognition and retrieval

- D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, November 2004. Available: <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision. Washington, DC, USA: IEEE Computer Society, 2003. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1238663
- O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," in IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, 2007, pp. 1-8. Available: <http://dx.doi.org/10.1109/ICCV.2007.4408891>
- N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3d," ACM Transactions on Graphics (TOG), (SIGGRAPH) 2006. <http://phototour.cs.washington.edu/>

Remember: one page summary describing main results in each paper and how readings relate to each other due by email (to trevor@eecs.berkeley.edu) before start of class. One page total for all readings each week, not one page per paper.