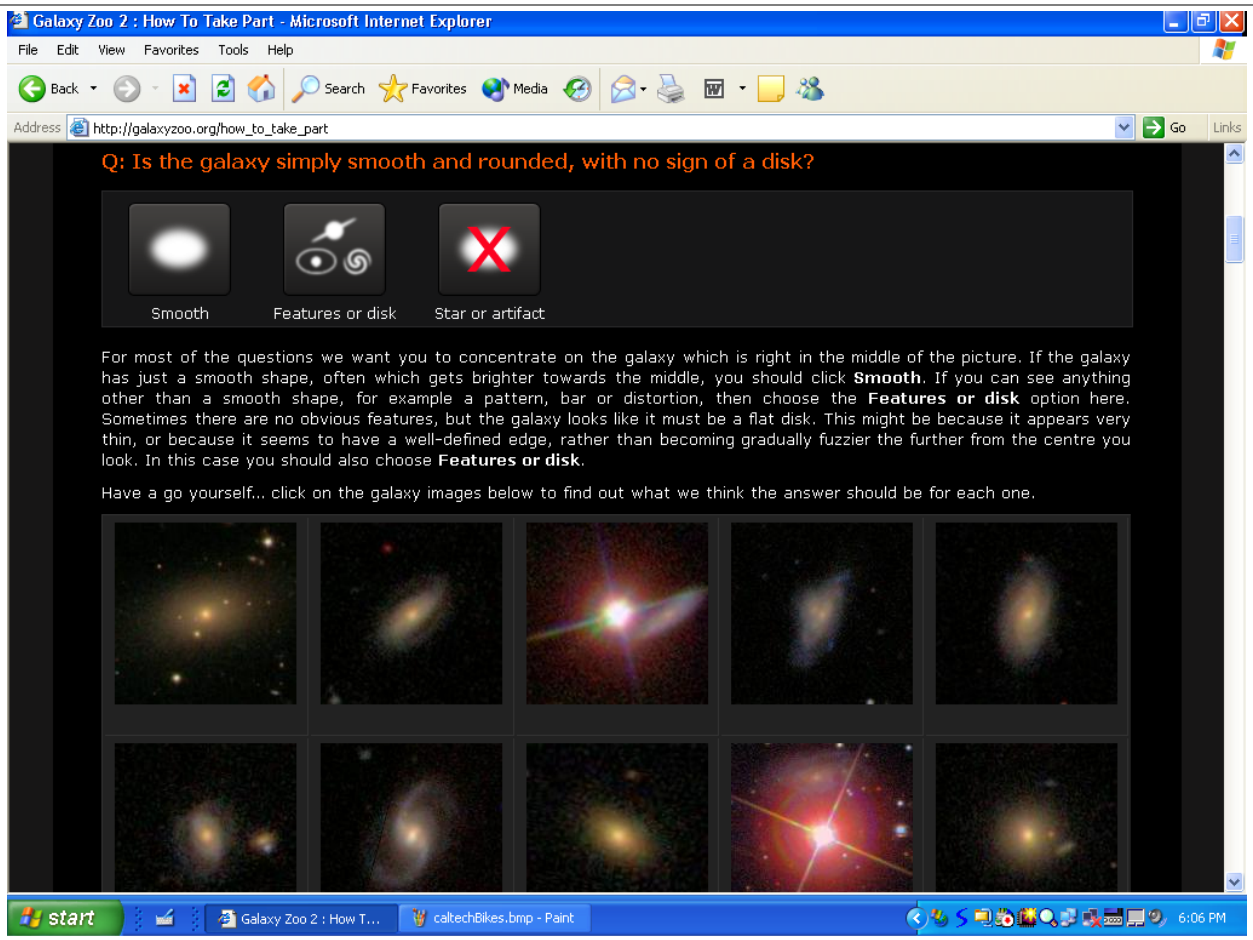
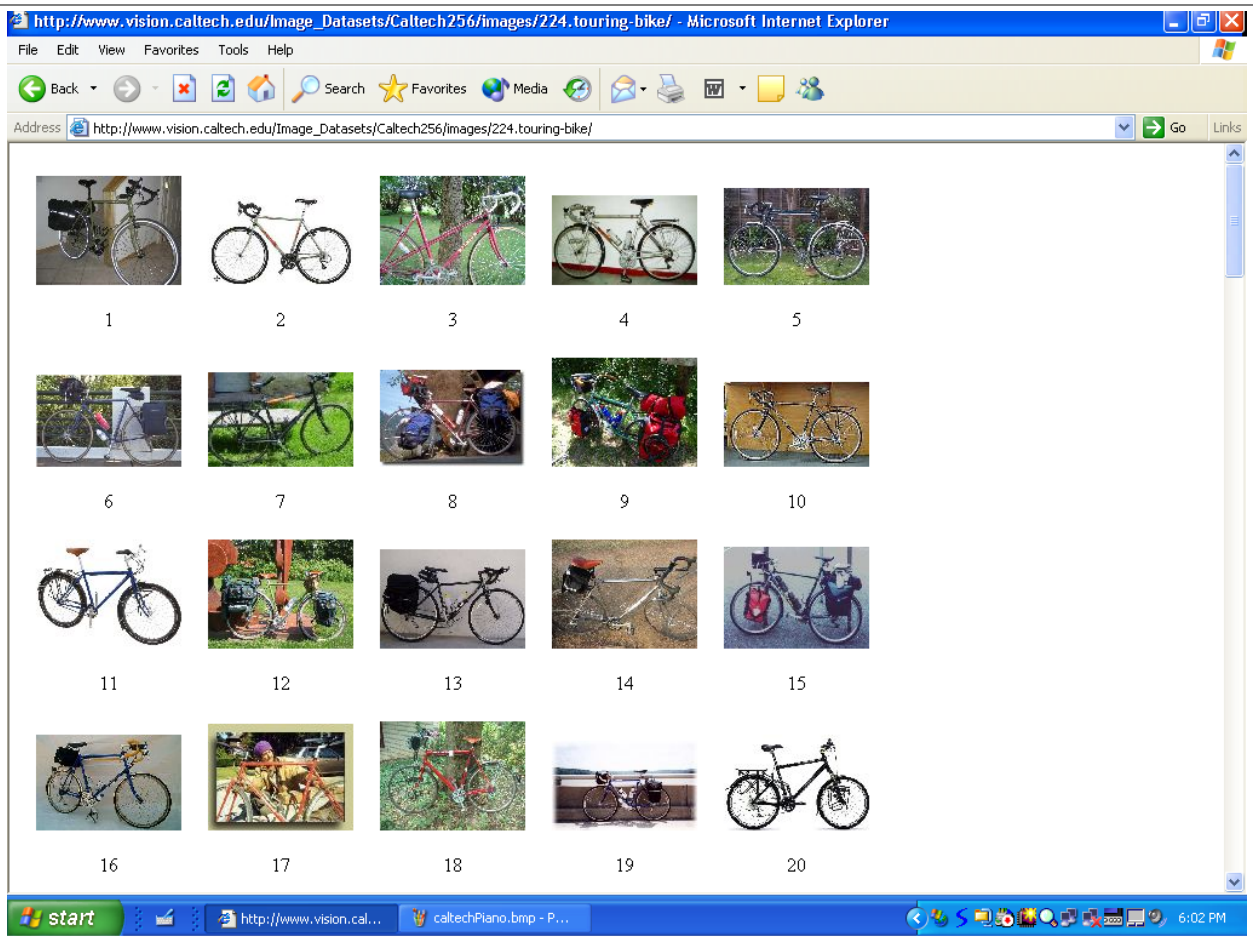
Could existing computer vision tools assist/replace humans in the Galaxy Zoo?

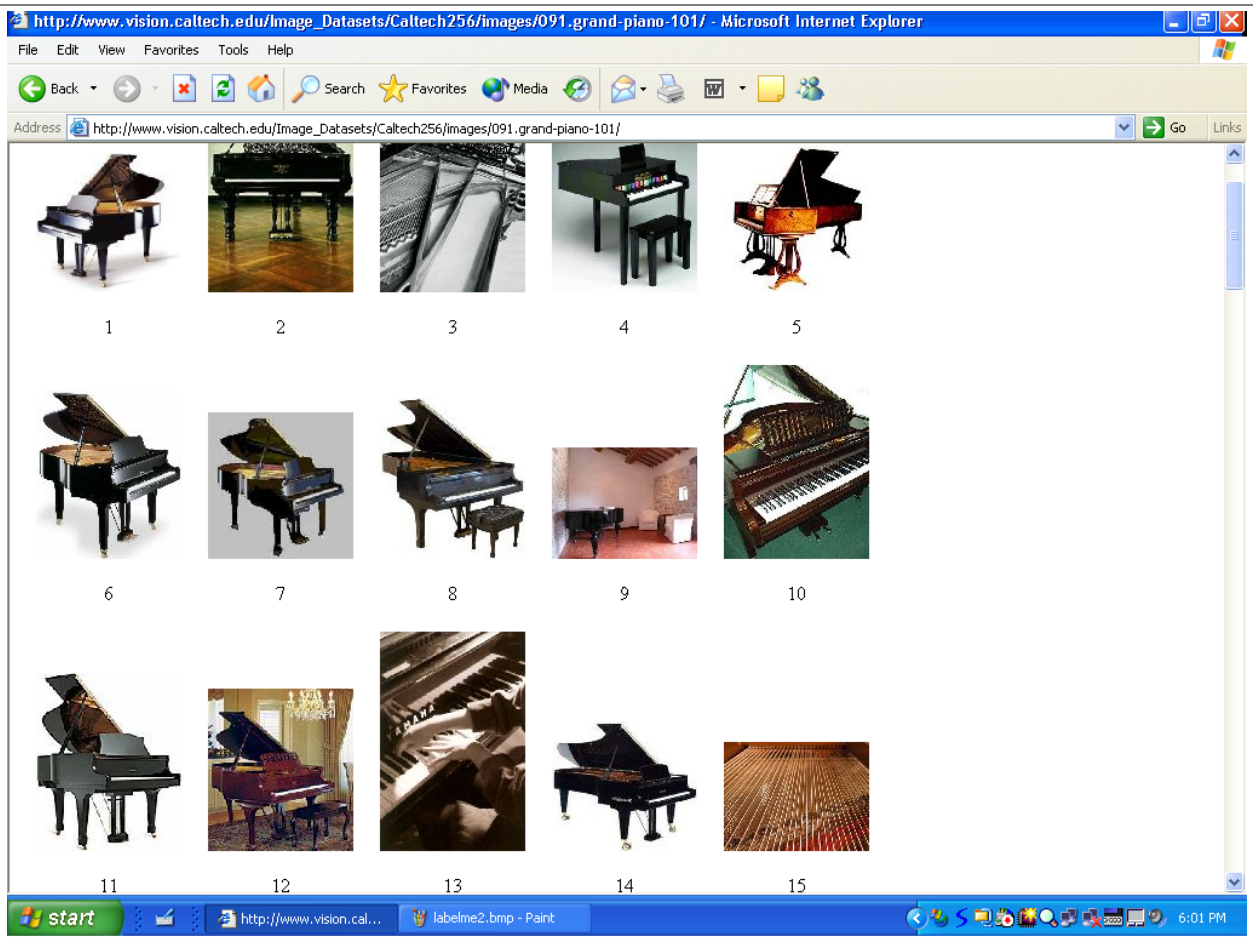
Nebojsa Jojic, Alessandro Perina

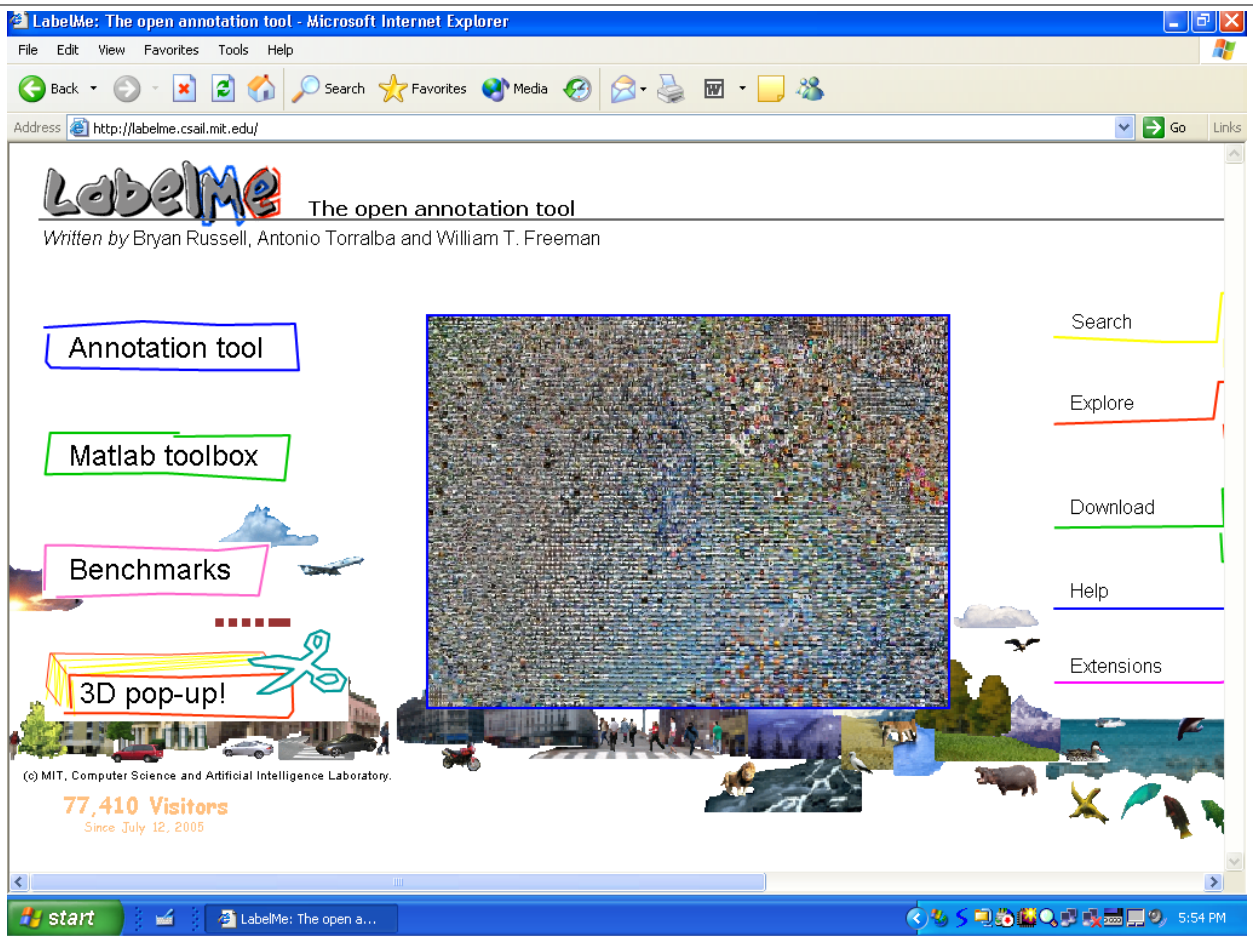
Microsoft Research

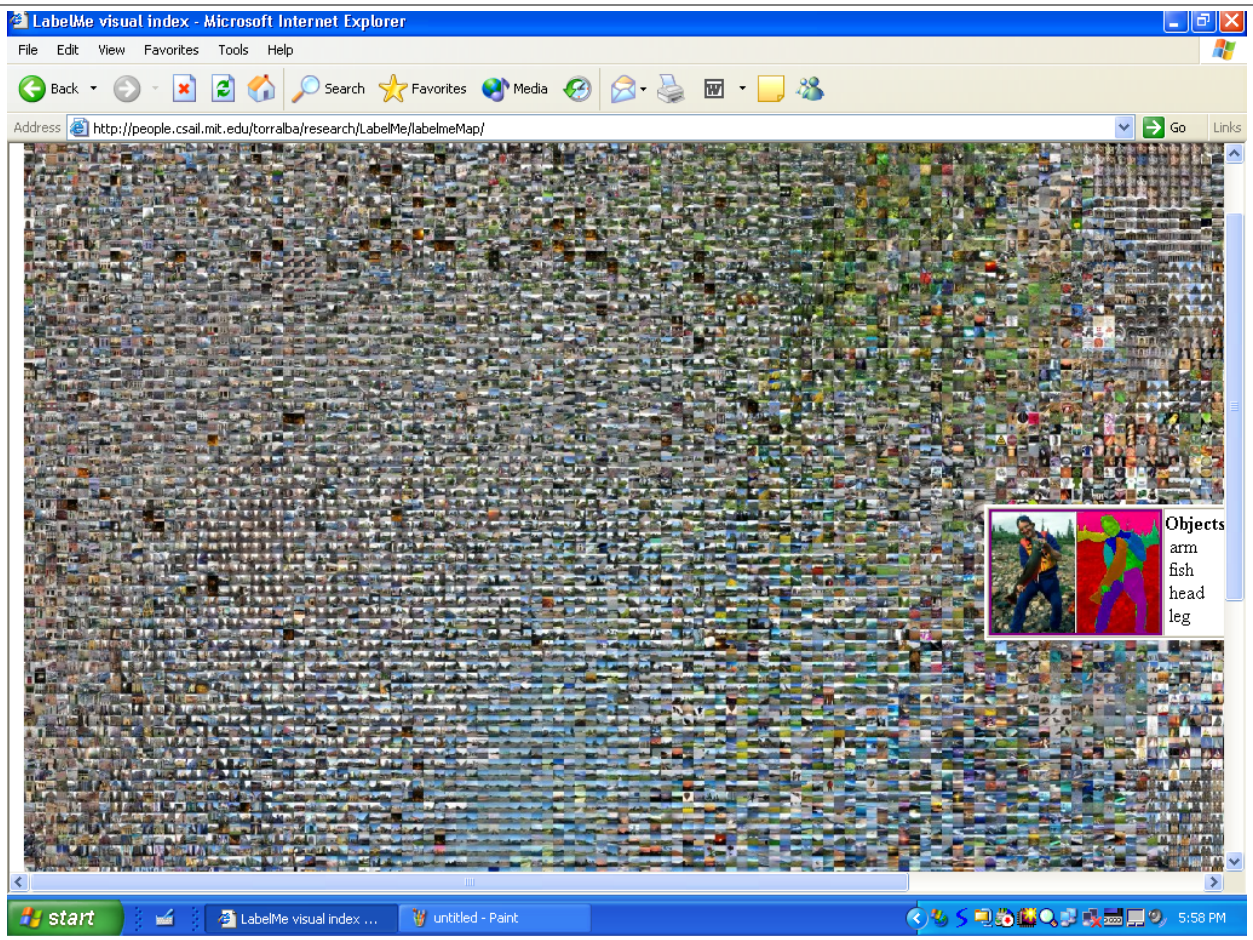
jojic@microsoft.com




http://www.vision.caltech.edu/Image_Datasets/Caltech256/images/224.touring-bike/ - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.vision.caltech.edu/Image_Datasets/Caltech256/images/224.touring-bike/
Go Links
1 2 3 4 5
6 7 8 9 10
11 12 13 14 15
16 17 18 19 20
start
http://www.vision.cal...
caltechPiano.bmp - P...
6:02 PM


http://www.vision.caltech.edu/Image_Datasets/Caltech256/images/091.grand-piano-101/ - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.vision.caltech.edu/Image_Datasets/Caltech256/images/091.grand-piano-101/
Go Links
1 2 3 4 5
6 7 8 9 10
11 12 13 14 15
start
http://www.vision.cal...
labelme2.bmp - Paint
6:01 PM


LabelMe: The open annotation tool - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://labelme.csail.mit.edu/
Go Links
LabelMe
The open annotation tool
Written by Bryan Russell, Antonio Torralba and William T. Freeman
Annotation tool
Matlab toolbox
Benchmarks
3D pop-up!
Search
Explore
Download
Help
Extensions
(c) MIT, Computer Science and Artificial Intelligence Laboratory.
77,410 Visitors
Since July 12, 2005
start
LabelMe: The open a...
5:54 PM


LabelMe visual index - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://people.csail.mit.edu/torralba/research/LabelMe/labelmeMap/
Go Links
Objects
arm
fish
head
leg
start
LabelMe visual index ...
untitled - Paint
5:58 PM

SIFT features (Lowe, IJCV 2004)

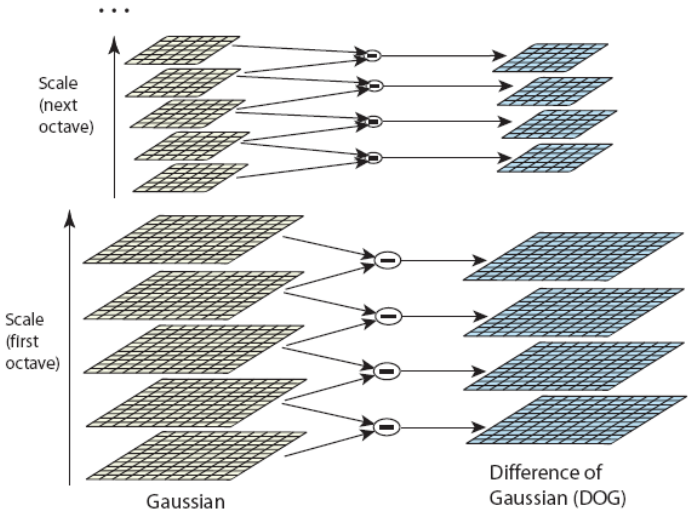
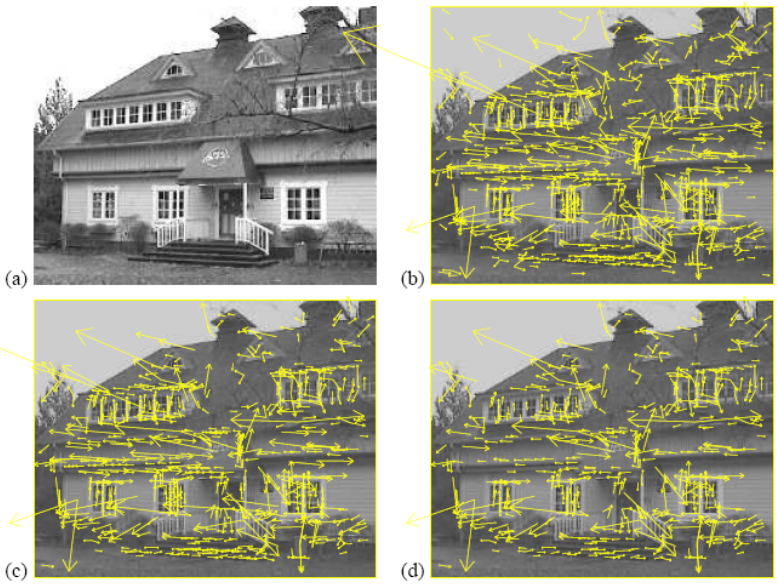


Figure 1: For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale space images shown on the left. Adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images on the right. After each octave, the Gaussian image is down-sampled by a factor of 2, and the process repeated.

SIFT features



SIFT features

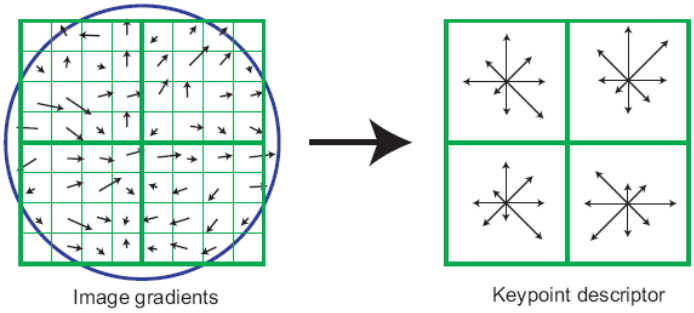


Figure 7: A keypoint descriptor is created by first computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location, as shown on the left. These are weighted by a Gaussian window, indicated by the overlaid circle. These samples are then accumulated into orientation histograms summarizing the contents over 4x4 subregions, as shown on the right, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. This figure shows a 2x2 descriptor array computed from an 8x8 set of samples, whereas the experiments in this paper use 4x4 descriptors computed from a 16x16 sample array.

SIFT features

SIFT features

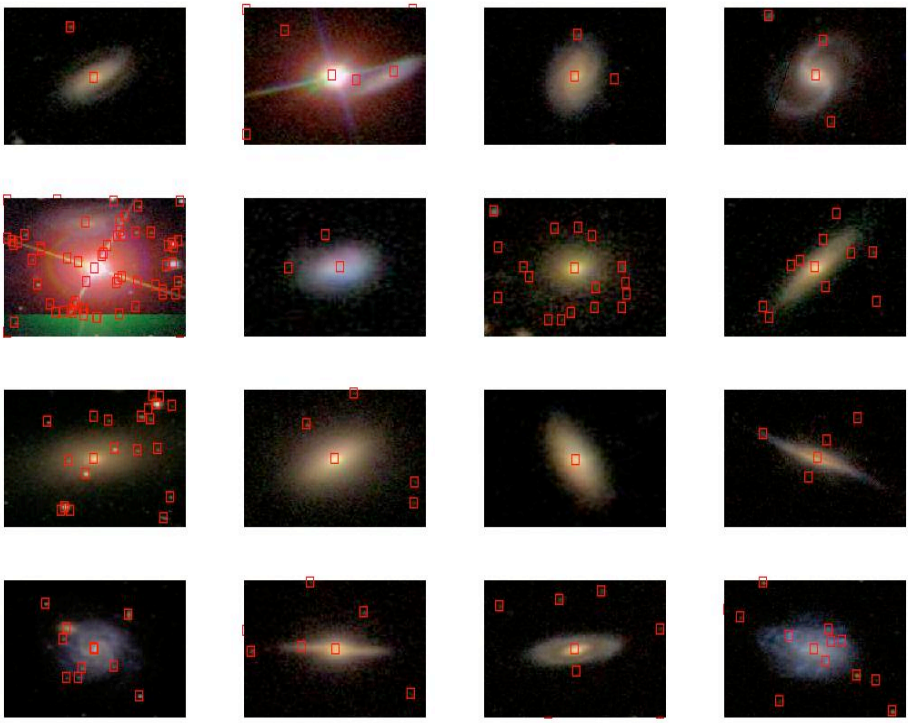
Figure 12: The training images for two objects are shown on the left. These can be recognized in a cluttered image with extensive occlusion, shown in the middle. The results of recognition are shown on the right. A parallelogram is drawn around each recognized object showing the boundaries of the original training image under the affine transformation solved for during recognition. Smaller squares indicate the keypoints that were used for recognition.

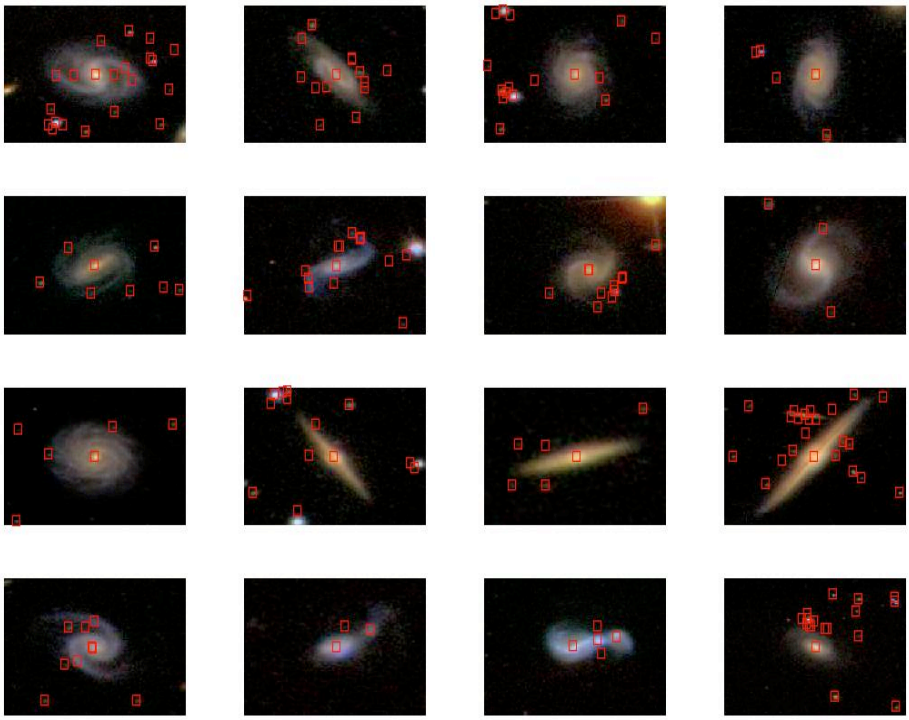
Distributions over features

- Trees/Forests
- Multinomial PCA models
- Undirected graphs
- …

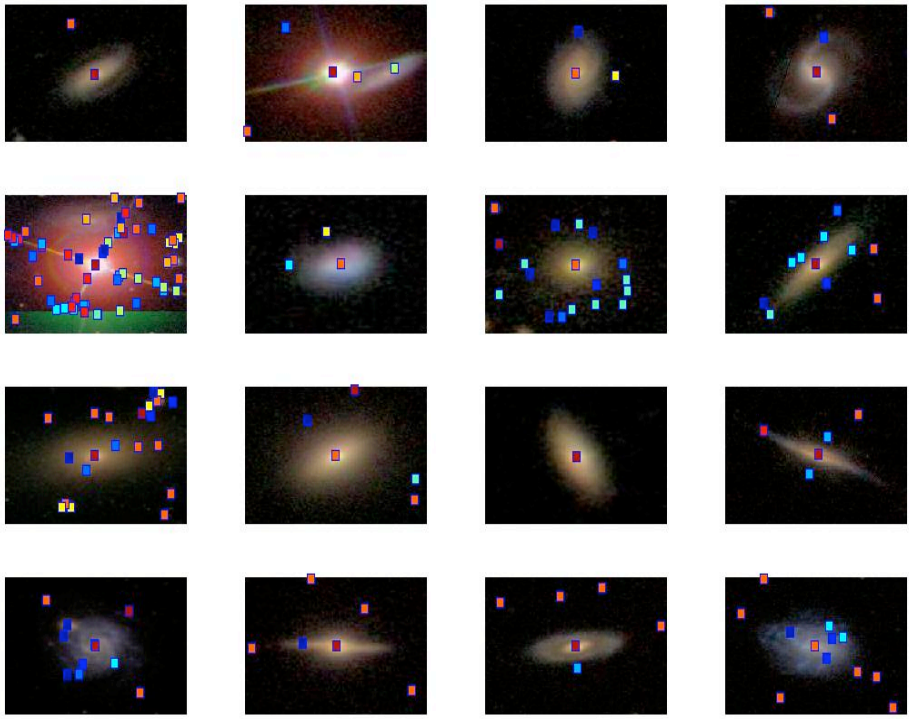
- Extensively tried out on Caltech databases

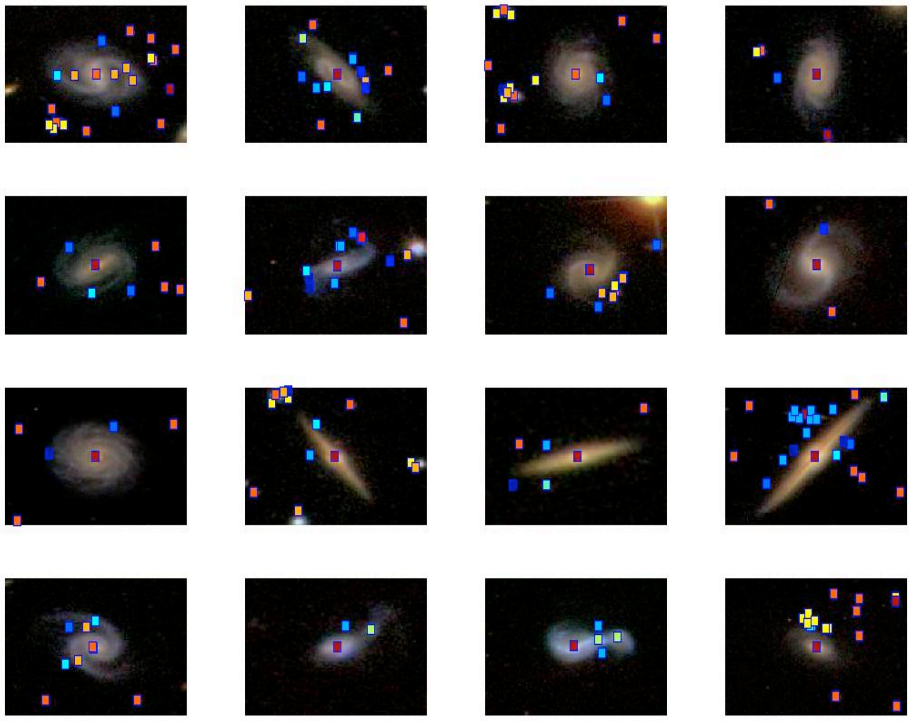
SIFT features in Galaxy Zoo images



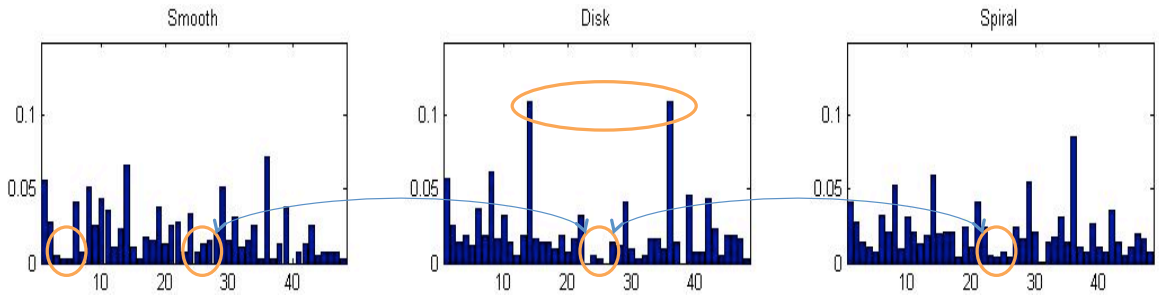


- The 128-dim SIFT descriptors are clustered to create the visual vocabulary (Cluster centers)





- Using a large vocabulary some class-specific codewords seem emerge
  - 150 Images
  - 2150 Codewords



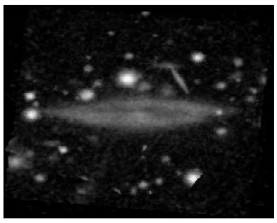
Learning image representations

[…]

Transformation

[…]

Mean and Standard Deviation



Epitomes

Input image

A set of image patches

Epitome

The idea

Input image

A set of image patches

Epitome

An epitome as a generative model of image patches

Epitome **e**:

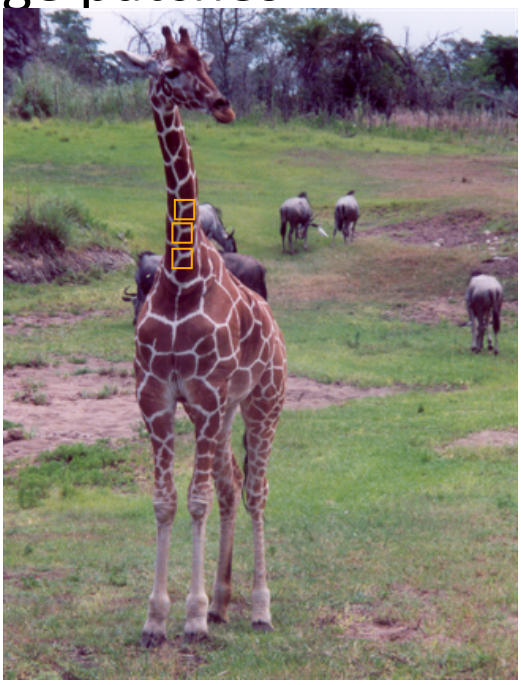
Mean μ

Variances ϕ

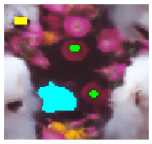
$$p(x_S \mid e_T) = \tilde{O}_k N(x_{S(k)}; m_{T(k)}, f_{T(k)})$$

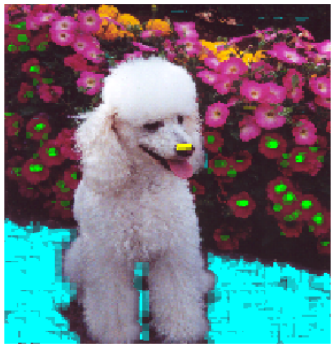
Prior probability of using an epitome patch T (or its average responsibility for the data)

$p(T)$ or $p(e_T)$



Compact representation



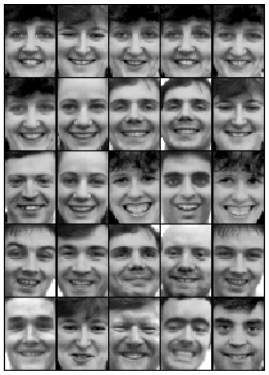


Compact representation

Training epitome on a set of images

The smiling point

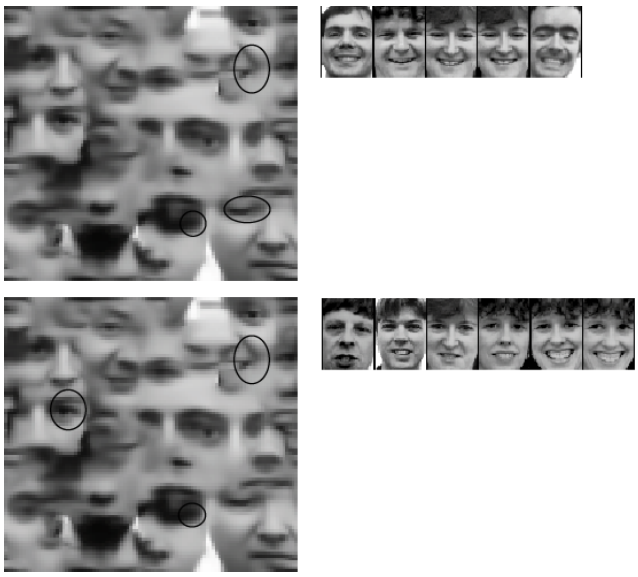
Epitome of 295 face images



Images with the highest total posterior at the "smiling point"

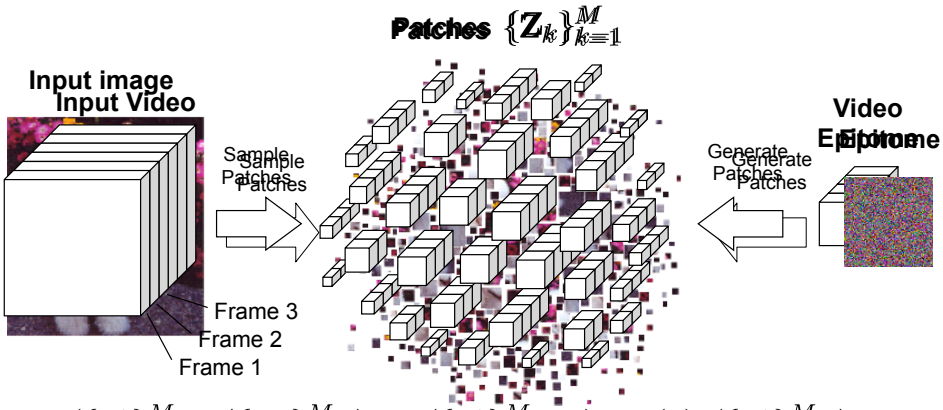
Images with the lowest total posterior at the "smiling point"

Multiple-cause pattern recognition



Galaxy Zoo epitomes

Video epitome



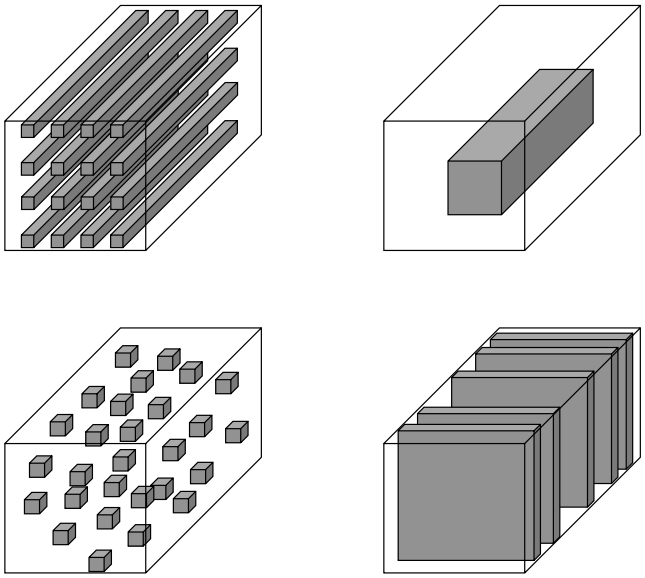
$$p(\{\mathcal{T}_k\}_{k=1}^{M}, \mathbf{e}|\{\mathbf{Z}_k\}_{k=1}^{M}) \approx q(\{\mathcal{T}_k\}_{k=1}^{M}, \mathbf{e}) = q(\mathbf{e})q(\{\mathcal{T}_k\}_{k=1}^{M})$$

$$p(\{\boldsymbol{Z}_k, \mathcal{T}_k\}_{k=1}^{M}, \boldsymbol{e}) = p(\boldsymbol{e}) \prod_{k=1}^{M} p(\mathcal{T}_k) p(\boldsymbol{Z}_k|\mathcal{T}_k, \boldsymbol{e}) = \delta(\mathbf{e} - \hat{\mathbf{e}}) \prod_{k=1}^{M} q(\mathcal{T}_k)$$

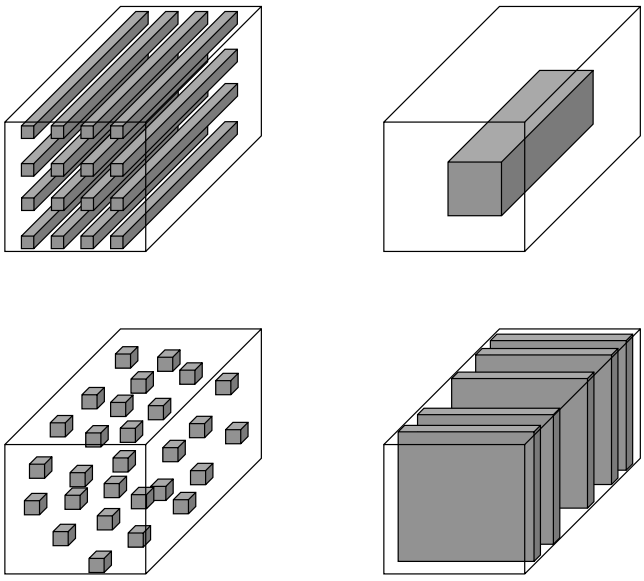
$$F(q,p) = \sum_{\{\mathcal{T}_k\}_{k=1}^{M}} \int_{\mathbf{e}} q(\{\mathcal{T}_k\}_{k=1}^{M}, \mathbf{e}) \log \frac{q(\{\mathcal{T}_k\}_{k=1}^{M}, \mathbf{e})}{p(\{\mathbf{Z}_k, \mathcal{T}_k\}_{k=1}^{M}, \mathbf{e})} \geq -\log p(\{\mathbf{Z}_k\}_{k=1}^{M})$$

$$p(\boldsymbol{Z}_k|\mathcal{T}_k, \boldsymbol{e}) = \prod_{i \in S_k} \mathcal{N}(\mathbf{z}_{k,i}; \mu_{\mathcal{T}_k(i)}, \phi_{\mathcal{T}_k(i)})$$

Missing Observations Scenarios



Missing Observations Scenarios



Missing data estimate from corrupted video alone

Computational efficiency

- Example: Viola-Jones face detector

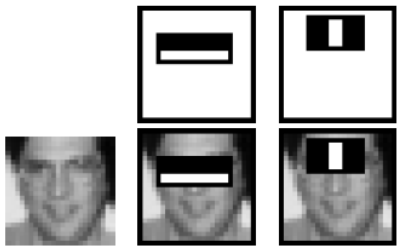
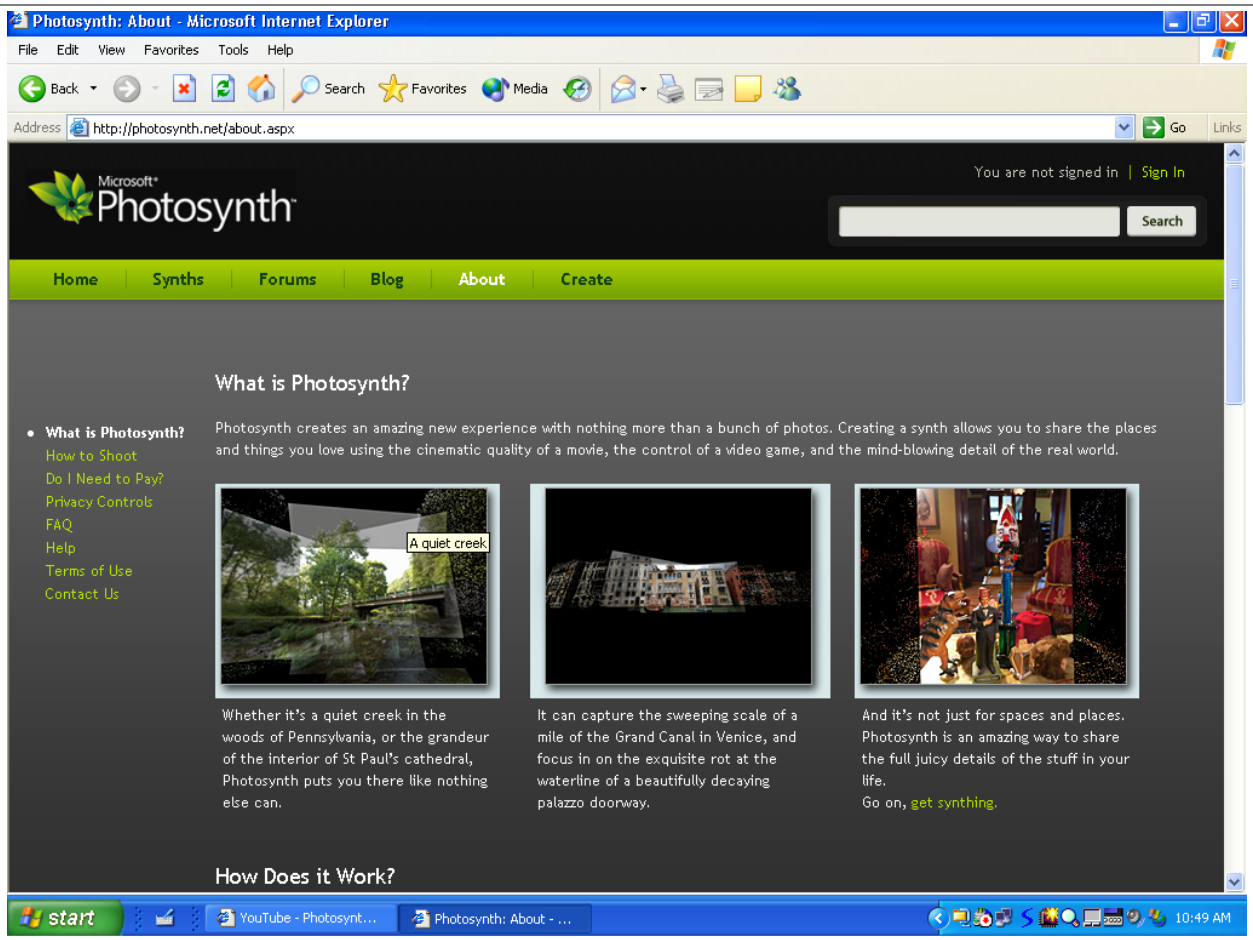
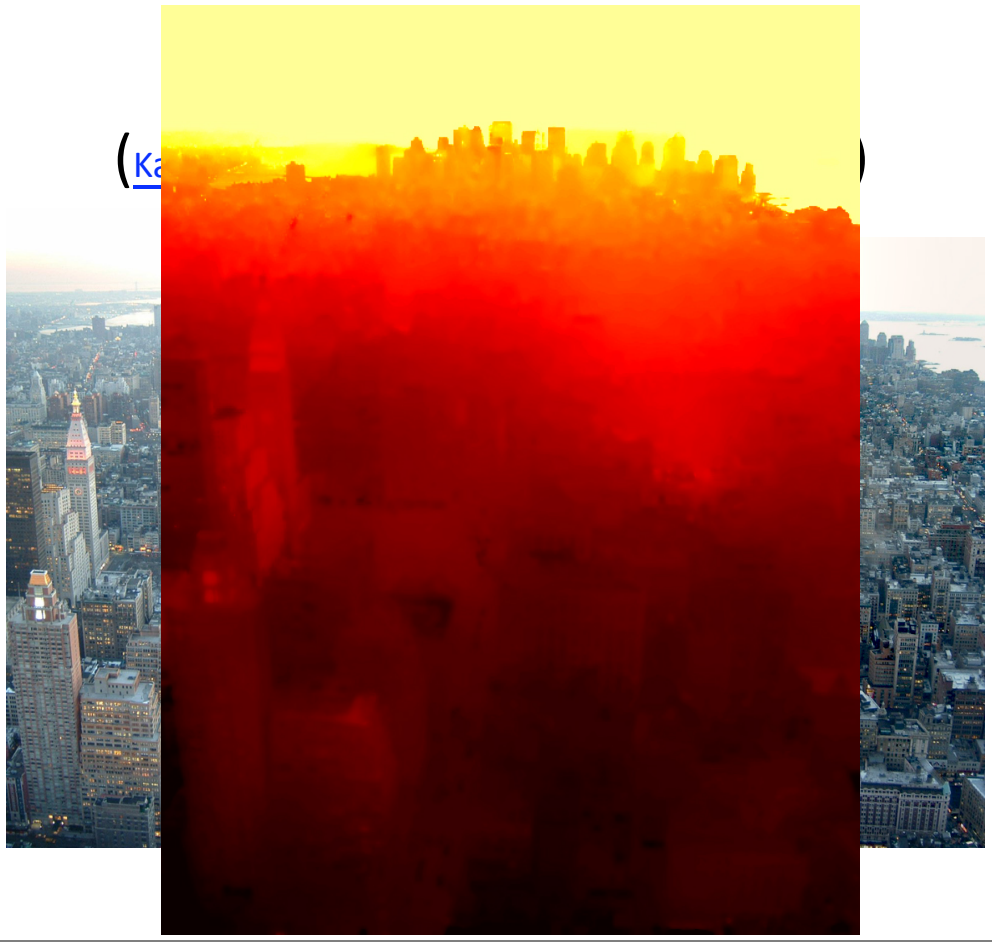


Figure 3: The first and second features selected by AdaBoost. The two features are shown in the top row and then overlayed on a typical training face in the bottom row. The first feature measures the difference in intensity between the region of the eyes and a region across the upper cheeks. The feature capitalizes on the observation that the eye region is often darker than the cheeks. The second feature compares the intensities in the eye regions to the intensity across the bridge of the nose.

The role of the image formation model


Photosynth: About - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://photosynth.net/about.aspx
Go Links
Microsoft Photosynth
You are not signed in | Sign In
Search
Home Synths Forums Blog About Create
What is Photosynth?
How to Shoot
Do I Need to Pay?
Privacy Controls
FAQ
Help
Terms of Use
Contact Us
What is Photosynth?
Photosynth creates an amazing new experience with nothing more than a bunch of photos. Creating a synth allows you to share the places and things you love using the cinematic quality of a movie, the control of a video game, and the mind-blowing detail of the real world.
A quiet creek
Whether it's a quiet creek in the woods of Pennsylvania, or the grandeur of the interior of St Paul's cathedral, Photosynth puts you there like nothing else can.
It can capture the sweeping scale of a mile of the Grand Canal in Venice, and focus in on the exquisite rot at the waterline of a beautifully decaying palazzo doorway.
And it's not just for spaces and places. Photosynth is an amazing way to share the full juicy details of the stuff in your life.
Go on, get synthing.
How Does it Work?
start
YouTube - Photosynt...
Photosynth: About - ...
10:49 AM



Fake object databases

- Greebles (Yale / Brown)

Galaxy Zoo Competition?

- CVPR 2010 Workshop?